**TRƯỜNG ĐẠI HỌC XÂY DỰNG**

**KHOA CÔNG NGHỆ THÔNG TIN  
------------o0o------------**

****

BÀI NGHIÊN CỨU

ĐỀ TÀI: Giải thuật di truyền và các ứng dụng

***GV hướng dẫn: Thầy Phạm Hồng Phong***

**Nhóm sinh viên thực hiện (Nhóm 4)***:*

Dương Công Sơn 64CS3 MSSV 167464

Phạm Hoàng Đăng Trung 64CS3 MSSV 1551664

Trịnh Ngọc Dương 64CS3 MSSV 39864

Hoàng Thị Thanh Tú 64CS3 MSSV 208764

Bùi Quốc Việt 64CS3 MSSV 222964

***Hà Nội, tháng 12 năm 2021.***

Mục lục

[Lời nói đầu 4](#_Toc92394096)

[Mở đầu 5](#_Toc92394097)

[Lý do chọn đề tài 5](#_Toc92394098)

[Mục tiêu của bài tập nghiên cứu 6](#_Toc92394099)

[Ý nghĩa khoa học và ý nghĩa thực tiễn 6](#_Toc92394100)

[Chương 1. Tổng quan về thuật toán di truyền và bài toán lập lịch 7](#_Toc92394101)

[1.1. Thuật toán di truyền 7](#_Toc92394102)

[1.1.1. Giới thiệu về thuật toán 7](#_Toc92394103)

[1.1.2. Cấu trúc của thuật toán di truyền 8](#_Toc92394104)

[1.1.3. Mã giả của thuật toán 10](#_Toc92394105)

[1.1.4. Các tham số, công thức của giải thuật 10](#_Toc92394106)

[1.1.5. Các thành phần của giải thuật di truyền. 11](#_Toc92394107)

[1.1. Bài toán lập lịch 13](#_Toc92394108)

[1.2.1. Tìm hiểu chung 13](#_Toc92394109)

[1.2.2. Mô tả bài toán lập lịch tổng quát 14](#_Toc92394110)

[1.2.3. Những giải pháp thuật toán đã được đề ra cho bài toán Lập Lịch Thời Khóa Biểu 15](#_Toc92394111)

[Chương 2. Cài đặt thuật toán di truyền và bài toán lập lịch thời khóa biểu 16](#_Toc92394112)

[2.1. Chọn mô hình cá thể 16](#_Toc92394113)

[2.2. Tạo quần thể ban đầu 17](#_Toc92394114)

[2.3. Độ thích nghi – Chọn cá thể 18](#_Toc92394115)

[2.4. Thuật toán lai ghép – đột biến 18](#_Toc92394116)

[2.5. Chọn điểm dừng thuật toán 19](#_Toc92394117)

[Chương 3. Thiết kế hệ thống lập lịch thời khóa biểu 20](#_Toc92394118)

[3.1. Thiết kế cơ sở dữ liệu rang buộc của bài toán 20](#_Toc92394119)

[3.2. Các thành phần đối tượng của bài toán 21](#_Toc92394120)

[3.3. Thiết kế nhiễm sắc thể 21](#_Toc92394121)

[3.4. Các tham số của hệ thống và thao tác lai ghép 22](#_Toc92394122)

[3.4.1. Phép chọn cá thẻ vào lớp cha-mẹ 22](#_Toc92394123)

[3.4.1. Phép lai ghép 23](#_Toc92394124)

[3.4.2. Phép đột biến 24](#_Toc92394125)

[3.5. Độ thích nghi của hệ thống 25](#_Toc92394126)

[Chương 4. Các thực nghiệm của hệ thống 26](#_Toc92394127)

[Chương 5. Kết luận và đề xuất hướng phát triển. 28](#_Toc92394128)

[5.1. Kết quả đạt được 28](#_Toc92394129)

[5.2. Hạn chế - Hướng phát triển trong tương lai. 28](#_Toc92394130)

[Chương 6. Tài liệu tham khảo. 29](#_Toc92394131)

# Lời nói đầu

Trong môn học Phân tích và thiết kế thuật toán, chúng ta đã học qua những thuật toán nổi tiếng và những ứng dụng của các thuật toán trong cuộc sống đời thực. Tuy nhiên có những thuật toán vẫn chưa là giải pháp tối ưu cho một số bài toán, nên việc luôn phải nghiên cứu để tìm ra những thuật toán tốt nhất cho những bài toán trong cuộc sống.

Ngày nay, công nghệ thông tin phát triển về mọi lĩnh vực trong cuộc sống và giúp cho đời sống trở nên dễ dàng và tiện nghi hơn. Và trong giáo dục cũng vậy. Ở bài toán lần này, chúng ta sẽ đề xuất một phương pháp tính toán dựa trên quan sát về quá trình tiến hóa trong ự nhiên và ứng dụng vào bài toán lập lịch thời khóa biểu, công việc thường ngày trên trường mà các thầy cô vẫn đau đầu suy nghĩ.

Giải thuật di truyền đã thu hút được nhiều chú ý trong những năm gần đây. Và theo các nghiên cứu của các nhà khoa học, giải thuật này có sự vượt trội nổi bật so với những bài toán khác, khi áp dụng định luật của tự nhiên vào bài toán.

Để hoàn thành bài báo cáo đồ án kết thúc môn học lần này, nhóm chúng em xin chân thành cảm ơn thầy Phạm Hồng Phong đã nhiệt tình giúp đỡ chúng em trong suốt quá trình học môn “Phân tích và thiết kế thuật toán” để chúng em có thể hoàn thiện tốt đồ án lần này.

# Mở đầu

## Lý do chọn đề tài

Trong lĩnh vực Phân tích và thiết kế thuật toán, chúng ta luôn cố gắng để tìm ra được những thuật toán tối ưu nhất để áp dụng vào trong nhiều bài toán và trong đời sống thực tế. Việc tìm kiếm ra những thuật toán mới áp dụng vào đời sống, sẽ giúp tiết kiệm được thời gian cũng như nâng cao năng suất làm việc của con người. Chúng ta đã biết đến các thuật toán tìm kiếm không có thông tin/vét cạn (tìm kiếm trên danh sách, trên cây hoặc đồ thị ) sử dụng phương pháp đơn giản nhất và trực quan nhất hoặc các thuật toán tìm kiếm có thông tin sử dụng Heuristics để áp dụng các tri thức về cấu trúc của không gian tìm kiếm nhằm giảm thời gian cần thiết cho việc tìm kiếm được sử dụng nhiều nhưng chỉ với không gian tìm kiếm nhỏ và không hiệu quả khi tìm kiếm trong không gian tìm kiếm lớn. Tuy nhiên, trong thực tiễn có rất nhiều bài toán tối ưu với không gian tìm kiếm rất lớn cần phải giải quyết. Vì vậy, việc đòi hỏi thuật giải chất lượng cao và sử dụng kỹ thuật lập trình tốt rất cần thiết khi giải quyết các bài toán có không gian tìm kiếm lớn.

Trong chủ đề lần này, chúng tôi nhận thấy được sự phát triển của công nghệ thông tin trong nhiều lĩnh vực và chúng tôi đã tìm hiểu đưa công nghệ thông tin vào giáo dục, cụ thể là công việc lập lịch khó khăn hằng ngày của ngành giáo dục. Chúng tôi cũng đã nghiên cứu và đưa ra thuật toán khá mới và giải quyết rất tốt bài toán này là thuật toán di truyền (Genetic Algorithm)

Lập lịch là một trong những chủ đề quan trọng thuộc lĩnh vực vận trù học xuất hiện từ đầu năm 1950. Mục tiêu chính của lập lịch là phân phối tài nguyên dùng chung một cách hiệu quả nhất cho các tác vụ đồng thời trong toàn bộ nhớ thời gian xử lý. Một mô hình chung nhất về lập lịch đó là bài toán lập lịch thời khóa biểu ( Scheduling Problem - SP). Bài toán lập lịch thời khóa biểu là một trong những bài toán được nghiên cứu nhiều nhất và là một mô hình phát triển tốt nhất về lý thuyết lập lịch. Nó cũng có một ứng dụng vô cùng mạnh mẽ trong thực tiễn, cuộc sống và sản xuất.

Trong các nghiên cứu đã có nhiều giải pháp được đề xuất cho bài toán lập lịch thời khóa biểu. Tuy nhiên, cho tới nay chưa có một thuật toán nào được đề xuất mà giải quyết được triệt để bài toán này. Ta có thể nêu ra một số vấn đề mà những giải pháp trên còn tồn tại như sau:

1. Xét về mặt dữ liệu, một số thuật toán đã đề xuất không thể linh hoạt trong việc sử dụng dữ liệu time series.
2. Các thuật toán đã đưa ra cho bài toán này đều chưa có chứng thực gì để có thể đánh giá được tính đúng đắn của thuật toán

Cho đến thời điểm hiện tại, ở nước ta việc nghiên cứu về bài toán lập lịch vẫn còn chưa thực sự phát triển, do vậy nhóm em lựa chọn đề tài: Lập lịch thời khóa biểu sử dụng giải thuật di truyền để nghiên cứu và giải quyết các vấn đề còn tồn đọng của bài toán này.

## Mục tiêu của bài tập nghiên cứu

Trong bài tập lớn này, nhóm em tập trung nghiên cứu và triển khai một số vấn đề chủ yếu sau đây:

1. Tìm hiểu về thuật toán di truyền, đánh giá với các thuật toán khác, chỉ ra những vấn đề còn tồn tại.
2. Áp dụng thuật toán di truyền vào bài toán Lập Lịch Thời Khóa Biểu.
3. Triển khai thuật toán di truyền
4. Xây dựng thành một ứng dụng giúp người dùng dễ dàng tương tác.

## Ý nghĩa khoa học và ý nghĩa thực tiễn

* **Ý nghĩa khoa học**

1. Nghiên cứu về tổng quan của thuật toán:

* Phân tích thuật toán, đưa ra các so sánh của thuật toán này với các thuật toán khác, chứng minh tính đúng đắn và độ phức tạp của thuật toán.
* Đánh giá các giải pháp thuật toán khác trong bài toán lập lịch thời khóa biểu. Trên cơ sở đó tìm ra một hướng phát triển để nghiên cứu về giải thuật Gas.

1. Nghiên cứu và xây dựng dữ liệu cho một bài toán thực tế dựa trên thuật toán di truyền học.

* **Ý nghĩa thực tiễn**

1. Bài tập lớn này ứng dụng trực tiếp vào bài toán lập lịch thời khóa biểu của các trường học hiện nay.
2. Các bạn sinh viên khác có thể tham khảo cách tiếp cận mới này của thuật toán GAs.

- Bố cục của BTL

Bài tập lớn này gồm … trang với … hình vẽ, …. bảng, … tài liệu tham khảo, ngoài phần mở đầu và kết luận, bài tập lớn được chia thành 5 chương như sau:

Chương 1: Tổng quan về thuật toán di truyền.

Chương 2: Bài toán lập lịch.

Chương 3: Cài đặt thuật toán di truyền vào bài toán xếp thời khóa biểu.

Chương 4: Thiết kế hệ thống lập lịch thời khóa biểu.

Chương 5: Các thực nghiệm thực tế.

Chương 6: Kết luận và đề xuất hướng phát triển.

# Chương 1. Tổng quan về thuật toán di truyền

## 1.1. Thuật toán di truyền

### 1.1.1. Giới thiệu về thuật toán

Trước hết, chúng ta hiểu về các khái niệm Di truyền và Tiến hóa là gì?

“Di truyền” là hiện tượng những lớp con cái được thừa hưởng những tính trạng của cha mẹ thông qua gen di truyền từ cha mẹ. Trong sinh học, di truyền sẽ chuyển những đặc trưng sinh học từ cha mẹ đến con cái, nó đồng nghĩa với việc di chuyển gen và gen được di chuyển mang thông tin di truyền.

“Tiến hóa” là khái niệm để nói đến quá trình biến đổi để hoàn thiện hơn các bộ phận, chức năng của sinh vật để phù hợp với điều kiện sinh tồn. Trong sinh học thì tiến hóa là sự thay đổi đặc tính di truyền của một quần thể sinh học qua những thế hệ nối tiếp nhau. Các quá trình tiến hóa đã làm đa dạng hóa mọi mức độ tổ chức sinh học bao gồm loài, các cá thể sinh vật và các phân tử như ADN và Protein. Tiến hóa do chọn lọc tự nhiên để thích nghi với môi trường nên ta có thể suy ra:

* Nhiều cá thể con được sinh ra hơn số lượng có thể sống sót
* Các tính trạng khác nhau giữa các cá thể, dẫn tới tỉ lệ sinh tồn và sinh sản khác nhau
* Những khác biệt về đặc điểm đều di truyền từ bố mẹ

Quá trình chọn lọc tự nhiên này bảo tồn được những đặc điểm được cho là phù hợp hơn với chức năng mà chúng đảm nhiệm. Vào đầu thế kỷ 20, di truyền học kết hợp với lý thuyết tiến hóa nhờ chọn lọc tự nhiên của Darwin thông qua di truyền học quần thể. Và là bước tiến cho những nghiên cứu sau này.

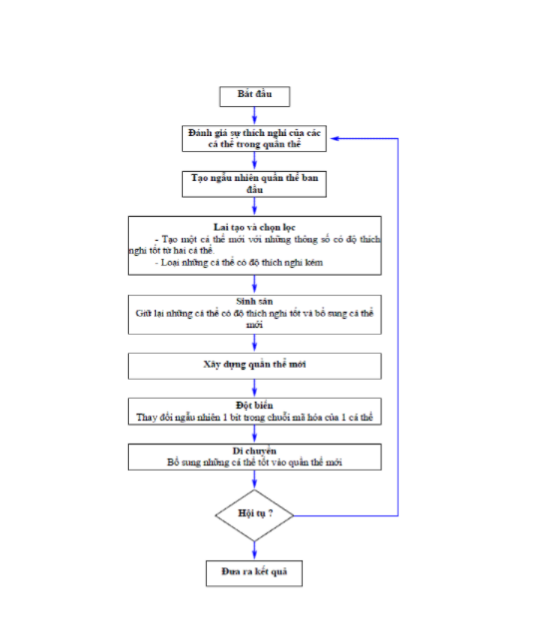
Thuật toán di truyền (Genetic Algorithms) là thuật toán được mô phỏng theo quá trình thích nghi tiến hóa, chọn lọc tự nhiên, là kĩ thuật chung giúp giải quyết vấn đề bài toán bằng cách mô phỏng sự tiến hóa của con người hay của sinh vật nói chung (dựa trên thuyết tiến hóa của Darwin hay luật di truyền Mendel). GAs thuộc lớp các giải thuật xuất sắc nhưng lại rất khác các giải thuật tìm kiếm ngẫu nhiên vì chúng kết hợp các phần tử tìm kiếm trực tiếp và ngẫu nhiên.  Khác biệt quan trọng giữa tìm kiếm của GA và các phương pháp tìm kiếm khác là GA duy trì và xử lý một tập các lời giải, gọi là một quần thể (population). Trong GA, việc tìm kiếm giả thuyết thích hợp được bắt đầu với một quần thể, hay một tập hợp có chọn lọc ban đầu của các giả thuyết. Các cá thể của quần thể hiện tại khởi nguồn cho quần thể thế hệ kế tiếp bằng các hoạt động lai ghép và đột biến ngẫu nhiên – được lấy mẫu sau các quá trình tiến hóa sinh học. Ở mỗi bước, các giả thuyết trong quần thể hiện tại được ước lượng liên hệ với đại lượng thích nghi, với các giả thuyết phù hợp nhất được chọn theo xác suất là các hạt giống cho việc sản sinh thế hệ kế tiếp, gọi là cá thể (individual). Cá thể nào phát triển hơn, thích ứng hơn với môi trường sẽ tồn tại và ngược lại sẽ bị đào thải. GA có thể dò tìm thế hệ mới có độ thích nghi tốt hơn. GA giải quyết các bài toán quy hoạch toán học thông qua các quá trình cơ bản: lai tạo (crossover), đột biến (mutation) và chọn lọc (selection) cho các cá thể trong quần thể. Dùng GA đòi hỏi phải xác định được: khởi tạo quần thể ban đầu, hàm đánh giá các lời giải theo mức độ thích nghi – hàm mục tiêu, các toán tử di truyền tạo hàm sinh sản.

//Ý tưởng từ quá trình tiến hóa tự nhiên, xuất phát từ một lớp các lời giải tiềm năng ban đầu, GA tiến hành tìm kiếm trên không gian lời giải bằng cách xây dựng lớp lời giải mới tốt hơn (tối ưu hơn) lời giải cũ. Quá trình xây dựng lớp lời giải mới được tiến hành dựa trên việc chọn lọc, lai ghép, đột biến từ lớp lời giải ban đầu. Quần thể lời giải trải qua quá trình tiến hoá: ở mỗi thế hệ lại tái sinh các lời giải tương đối tốt, trong khi các lời giải “xấu” thì chết đi.

Với mỗi bài toán, thuật toán GAs sẽ có những biến thể khác nhau, làm sao để có thể dễ dàng quy bài toán về chuỗi những giải pháp có thể của bài toán. Sau đó sử dụng phép chọn lọc, chọn ra giải pháp tốt nhất. Thuật toán di truyền gồm bốn quy luật cơ bản là: lai ghép, đột biến, sinh sản và chọn lọc tự nhiên.// //comment lại rồi anh thấy cái nào ok hơn thì lấy

### 1.1.2. Cấu trúc của thuật toán di truyền

Sơ đồ thuật toán tổng quát của GAs:



Quá trình lai ghép: Chúng ta sẽ thực hiện chọn ra một cặp cha mẹ, từ đó ghép cặp để tạo ra một nhiễm sắc thể mới mang đặc tính trội của cả cha và mẹ. Giả sử chuỗi nhiễm sắc thể của cha và mẹ đều có chiều dài là m. Tạo một số ngẫu nhiên k trong đoạn 1 đến m-1, hoán đổi các gen sau vị trí k cho nhau ta được hai chuỗi nhiễm sắc thể con lúc này là m11+m22 và m21+m12. Đưa 2 cá thể mới này tham gia vào quần thể và tiếp tục tham gia quá trình tiến hóa.

Quá trình đột biến: Trong quá trình lai ghép, có một số tính trạng con không được thừa hưởng từ hai chuỗi nhiễm sắc thể cha-mẹ thì sẽ xảy ra đột biến.

Quá trình sinh sản và chọn lọc:  Là quá trình các cá thể được sao chép dựa trên độ thích nghi của nó. Sau đó thực hiện chọn lọc và loại bỏ các cá thể xấu và để lại những cá thể tốt. Có một số tiêu chí sau khi thực hiện chọn lọc:

* Chọn lọc dựa trên độ thích nghi.
* Chọn lọc dựa trên sự xếp hạng.
* Chọn lọc dựa trên sự cạnh tranh

### 1.1.3. Mã giả của thuật toán

B1: Khởi tạo và mã hóa một quần thể ngẫu nhiên của NST. Đó gọi là “quần thể hiện tại”

B2:  Đánh giá độ thích nghi của mỗi NST trong quần thể hiện tại.

B3: Tạo ra thế hệ trung gian, thông qua chọn lựa suy diễn các NST trong quần thể hiện tại tuỳ theo độ thích nghi. Đó sẽ là cha mẹ của những thế hệ tiếp theo.

B4: Áp dụng toán tử lai ghép và đột biến đối với mỗi cặp NST đơn trong thế hệ trung gian, từ đó sinh ra một thế hệ NST mới. Đó là quần thể hiện tại.

Lặp lại các bước 2-4 cho đến khi một giải pháp phù hợp được tìm thấy.

|  |  |
| --- | --- |
| Bắt đầu:  t = 0;  Khởi tạo P(t)  Tính độ thích nghi cho các cá thể thuộc P(t)  Tiến hành lặp:     t = t + 1     Lựa chọn P(t)     Lai ghép P(t)     Đột biến P(t)  hết lặp  kết thúc |  |

### 1.1.4. Các tham số, công thức của giải thuật

* **Các tham số**

Xác suất lai ghép là tham số cho biết tần suất thực hiện toán tử lai ghép. Nếu không có lai ghép, cá thể con sẽ chính là bản sao của cá thể “cha mẹ”. Nếu xác suất lai ghép bằng 100%, khi đó mọi cá thể con đều đƣợc tạo ra qua quá trình lai ghép.

Xác suất đột biến (mutation rate): là tham số cho biết tần suất đột biến của nhiễm sắc thể. Nếu không có đột biến, thế hệ con đƣợc tạo ra ngay sau giai đoạn lai ghép mà không bị thay đổi. ngược lại, một hoặc một số phần của nhiễm sắc thể sẽ bị thay đổi. Nếu xác suất đột biến là 100%, toàn bộ nhiễm sắc thể sẽ bị thay đổi. Nếu tham số này bằng 0%, không có gì bị thay đổi hết

Kích thước quần thể (population size):  là tham số cho biết có bao nhiêu cá thể (NST) trong 1 thế hệ của quần thể. Nếu có quá ít cá thể, khả năng thực hiện lai ghép rất nhỏ và khi đó chỉ có một vùng tìm kiếm nhỏ mới được khảo sát. Ngược lại, việc kích thước quần thể quá lớn cũng không tốt, do nó sẽ làm chậm quá trình giải bài toán

* **Một số công thức**

Tính độ thích nghi Fitness của mỗi nhiễm sắc thể.

Độ thích nghi =

Ta có thể thấy khi số conflict là 0 thì độ thích nghi là 1(tương đương với 100%) và số conflict càng lớn thì độ thích nghi càng thấp. Độ thích nghi này có thể có công thức khác sau tùy thuộc vào từng bài toán.

Tiến hành chọn lọc bằng cách chọn ra một nhiễm sắc thể từ quần thể hiện hành vào quần thể mới sau khi chọn ngẫu nhiên r trong khoảng [0, 1] nếu r > 0.5 thì chọn nhiễm sắc thể 1 và ngược lại chọn nhiễm sắc thể 2.

### 1.1.5. Các thành phần của giải thuật di truyền.

* Tạo quần thể ban đầu

Tùy thuộc vào mỗi bài toán ta lựa chọn quần thể bắt đầu khác nhau. Khi áp dụng thuật toán di truyền vào mỗi bài toán ta cần xác định rõ nhiễm sắc thể và cá thể cho từng bài toán và thông thường nó sẽ là kết quả cuối cùng. Việc phân tích sẽ dựa trên kết quả cơ bản nhất.

* Đánh giá độ thích nghi

Thuật toán sẽ thực hiện kế thừa những cá thể có độ thích nghi cao, do vậy độ thích nghi sẽ được tính dựa trên số conflict mà nó mắc phải. Theo lý thuyết Darwin, nhiễm sắc thể tốt nhất sẽ tồn tại và tạo ra các cá thể con mới. Có nhiều phương pháp để chọn nhiễm sắc thể tốt nhất như:

* Chọn lọc Roulette (Roulette Wheel Selection)
* Chọn lọc xếp hạng (Rank Selection)
* Chọn lọc cạnh tranh (Tournament Selection)
* Toán tử lai ghép

Toán tử này giúp nâng cao kết quả của cá thể, do đó, toán tử lai ghép sẽ tạo điều kiện cho tiến trình hội tụ nhanh hay chậm. Có một số phương pháp lai ghép như sau:

* Lai ghép ánh xạ tường phần (PMX partial Mapped Crossover)
* Lai ghép có trật tự (OX order Crossover)
* Lai ghép dựa trên vị trí (Position Based Crossover)
* Lai ghép có chu trình (CX cycle crossover)
* Lai ghép thứ tự tuyến tính (LOX Linear Order Crossover)
* Toán tử đột biến

Tương tự như phép lai, toán tử đột biến cũng giúp tăng quá trình hội tụ, nhưng sẽ tăng một cách đột ngột, có một số phương pháp để đánh giá mức độ tốt của phương pháp đột biến như:

* Đột biến đảo ngược (Inversion Mutation)
* Đột biến chèn (Insertion Mutation)
* Đột biến thay thế (Displacement Mutation)
* Đột biến tương hỗ (Reciprocal Exchange Mutation)
* Đột biến chuyển dịch (Shift Mutation)
* Điều kiện kết thúc:
* Khi mức độ thích nghi của chúng ta đạt ở mức yêu cầu
* Số thế hệ vượt quá mức quy định trước
* Thời gian thực hiện vượt mức cho phép

## So sánh Gas với kỹ thuật tối ưu khác.

Hoạt động của GAs đơn giản là việc mô phỏng sự tiến hóa và chọn lọc tự nhiên bằng máy tính bắt đầu từ một quần thể ngẫu nhiên. Bên cạnh đó để tối ưu ta cần hàm lượng giá hoặc hàm thích nghi để chọn cá thể tốt và loại bỏ cá thể xấu. Thuật toán di truyền (GAs) khác với kĩ thuật tối ưu khác ở chỗ:

- GAs làm việc với bộ mã của biến chứ không phải làm việc trực tiếp trên biến.

- Hầu hết các kĩ thuật tối ưu thông thường tìm kiếm từ một đỉnh, trong khi đó GAs luôn hoạt động trên tập hợp đỉnh (điểm tối ưu), điều này là một ưu điểm của GAs giúp tăng cơ hội tiếp cận tối ưu toàn cục và tránh hội tụ sớm tại điểm cục bộ địa phương.

- GAs đánh giá hàm mục tiêu để phục vụ quá trình tìm kiếm, vì vậy có thể ứng dụng cho bất kì bài toán tối ưu nào (liên tục hay rời rạc).

- GAs thuộc lớp các thuật toán xác suất, các thao tác cơ bản của GAs dựa trên khả năng tích hợp ngẫu nhiên trong quá trình xử lý.

## Một số ứng dụng của GAs

Gas được sử dụng cho những bài toán khó, và đã được ứng dụng thành công cho một số bài toán như: Lập kế hoạch, điều khiển tương thích, chương trình trò chơi, các bài toán vận tại, bài toán người đi du lịch, ... Sau đây là một và ứng dụng tiêu biểu của Gas.

* **Bài toán người du lịch (TSP)**

TSP được mô tả như sau: Một du khách muốn thăm những thành phố anh quan tâm; mỗi thành phố thăm qua đúng một lần; rồi trở về điểm khởi hành. Biết trước chi phí di chuyển giữa hai thành phố bất kỳ. Yêu cầu của bài toán là xây dựng một lộ trình thỏa các điều kiện trên với tổng chi phí nhỏ nhất.

TSP là bài toán tối ưu tổ hợp, không gian tìm kiếm là tập các hoán vị của n thành phố. Bất cứ hoán vị nào của n thành phố cũng là một lời giải chấp nhận được. Lời giải tối ưu là một hoán vị với chi phí tối thiểu của hành trình. Không gian tìm kiếm là n!. Có thể giải bài toán này bằng nhiều phương pháp: phương pháp nhánh cận, phương pháp gần đúng hay những phương pháp tìm kiếm heuristic. Phương pháp nhánh cận đã được chứng minh đạt sự tối ưu về lời giải, tuy nhiên phương pháp này lại mất khá nhiều thời gian khi số đỉnh của đồ thị lớn.

Trong những năm gần đây, đã xuất hiện nhiều thuật toán đạt gần đến lời giải tối ưu của bài toán TSP: láng giềng gần nhất, đảo gần nhất, đảo xa nhất…và TSP cũng trở thành một đích ngắm của cộng đồng GAs. Với bài toán này chúng ta sẽ đánh số các thành phố và dùng một vector nguyên để biểu diễn một NST lộ trình v= biểu diễn một lộ trình: từ i1 đến i2…, từ in-1 đến in và trở về i1 (v là một hoán vị của vector ), hàm lượng giá chính là chi phí của lộ trình.

* **Bài toán lập lịch**

Lập lịch là bài toán tổ chức sản xuất. Một công ty cần sản xuất nhiều loại hàng hóa; những hàng hóa này có thể được sản xuất theo những kế hoạch khác nhau. Mỗi kế hoạch xử lý gồm một chuỗi thao tác; những thao tác này sử dụng một số tài nguyên và cần thời gian chạy máy. Một lịch sản xuất là một kế hoạch thực hiện các đơn đặt hàng. Trong đó, một số đơn đặt hàng có thể được thực hiện với cùng những thao tác tương đương. Nhiệm vụ là lên kế hoạch, lập lịch sản xuất, để thực hiện các đơn đặt hàng này nhanh nhất có thể.

Bài toán lập lịch là chọn một chuỗi các thao tác đồng thời chỉ định về thời gian bắt đầu/ kết thúc và các tài nguyên cần thiết cho mỗi thao tác. Điều cần quan tâm chính yếu là chi phí thời gian máy rỗi, năng lực lao động và các đơn đặt hàng cần hoàn thành đúng hạn.Ý tưởng chính trong phương pháp là mã hóa biểu diễn của lịch phân công là các toán tử di truyền phải thực hiện theo cách có ý nghĩa, và một bộ giải mã phải luôn tạo ra một lời giải hợp lệ cho bài toán. Thủ tục giải mã mô phỏng các thao tác của công việc theo cách mà khi một máy tính sẵn sàng chọn lựa, thì thao tác cho phép đầu tiên từ danh sách ưu tiên được lấy ra. Ví dụ nếu danh sách ưu tiên của máy m1 là: m1(40 o3 o1 o2 ‘chờ’ ‘nhàn rỗi’), thì thủ tục giải mã vào thời điểm 40 có thể thực hiện đơn đặt hàng o3 trên máy m1. Nếu không được, thủ tục giải mã sẽ thực hiện đơn đặt hàng o1 và o2 (nghĩa là, tìm ở o1 trước; nếu không được mới tìm ở o2). Biểu diễn này bảo đảm tạo một lịch phân công hợp lệ.

* **Bài toán lập thời khóa biểu cho trường học.**

Bài toán thời khóa biểu là một bài toán kết hợp nhiều ràng buộc không tầm thường thuộc nhiều loại. Có nhiều phiên bản của bài toán thời khóa biểu, một trong những bài toán này có thể được mô tả như sau: Có một danh sách các giáo viên, một danh sách các khoảng thời gian, một danh sách các lớp. Bài toán cần tìm thời khóa biểu tối ưu (giáo viên – thời gian – lớp); hàm mục tiêu phải thỏa những mục tiêu này (các ràng buộc mềm) gồm: Có một số giờ được xác định trước cho mỗi giáo viên và mỗi lớp; Chỉ một giáo viên trong một lớp vào một giờ nhất định; Một giáo viên không thể dạy hai lớp cùng lúc; Đối với mỗi lớp được xếp thời khóa biểu vào một khoảng thời gian, phải có một giáo viên…Ngoài ra còn có các mục tiêu sư phạm như trải một số lớp ra nguyên tuần, những mục tiêu thuộc cá nhân như những giáo viên hợp đồng không phải dạy buổi chiều, và các mục tiêu về tổ chức như mỗi giờ có một giáo viên bổ sung sẵn sàng chỗ dạy tạm thời.

Biểu diễn nhiễm sắc thể tự nhiên nhất cho bài toán này là biểu diễn ma trận: một ma trận  , ở đây mỗi hàng tương ứng với một giáo viên, mỗi cột tương ứng với một giờ, các phần tử của ma trận R là các lớp  . Các ràng buộc chủ yếu được xử lý bởi các toán tử di truyền và thuật giải sửa chữa được sử dụng để loại bỏ những trường hợp mà có nhiều hơn một giáo viên xuất hiện trong cùng một lớp vào cùng một giờ.

* **Phân hoạch đối tượng đồ thị**

Bài toán phân hoạch là cần chia n đối tượng thành k loại. Lớp bài toán này gồm nhiều bài toán nổi tiếng như bài toán đóng thùng (gán các mặt hàng vào các thùng), bài toán tô màu đồ thị (gán các nút của đồ thị vào các màu cụ thể…).

Bài toán đóng thùng (Bin Packing – BP) được phát biểu như sau: Cho danh sách gồm n đồ vật L=a1,a2,a3,…,an và các thùng giống nhau với cùng sức chứa B. Kích thước của đồ vật a\_i là s\_i thỏa mãn . Vấn đề đặt ra là tìm cách xếp các đồ vật vào các thùng sao cho số lượng thùng cần phải sử dụng là ít nhất.

* **Vạch đường cho robot di chuyển**

Tìm đường là hướng dẫn robot di chuyển đến đích mà không bị lạc hay va vào những đối tượng khác. Trong bài toán này, một lộ trình được lập trước để robot đi theo, lộ trình này có thể dẫn dắt robot đi tới đích một cách hoàn hảo. Tuy nhiên, các nhà khoa học muốn cải tiến hơn bằng cách vạch lộ trình nội tại, phụ thuộc vào tri thức thu được từ việc cảm nhận môi trường cục bộ để xử lý các chướng ngại chưa biết.

Bộ tìm đường tiến hóa (EN) được đề xuất. Phần đầu của thuật giải là tìm lộ trình toàn cục tối ưu từ điểm khởi đầu đến đến đích, phần thứ hai có trách nhiệm xử lý những va chạm có thể xảy ra hay những vật cản chưa được biết trước bằng cách thay một phần của lộ trình toàn cục gốc bằng một lộ trình con tối ưu.1.4.

## Một số công trình nghiên cứu.

Về mặt tổng quát, GAs là một thuật toán tốt, tuy nhiên nó không phải là thuật toán thành công nhất đối với một số miền đặc biệt. Do đó, người ta mong muốn tìm ra lời giải tốt nhất có thể bằng cách kết hợp GAs với một số thuật toán hoặc kĩ thuật khác. Bài báo về TSP: “Meta-Heuristic-Kết hợp thuật giải di truyền với thông tin thống kê xác suất giải quyết bài toán người du lịch”.

Phương pháp gồm hai bước: Bước 1, sử dụng GAs để tìm thông tin thống kê xác suất các cung sẽ xuất hiện trong chu trình tối ưu từ các cá thể chu trình tốt nhất chọn lọc qua các thế hệ. Bước 2, từ các thông tin tìm được, thực hiện lại thuật toán GAs để được chu trình kết quả tối ưu, trong đó các phép lai ghép sẽ dựa vào thông tin xác suất tìm được ở Bước 1.

Bài báo về bài toán lập lịch: “Một thuật toán di truyền hiệu quả cho bài toán lập lịch job shop” . Bài báo này trình bày một thuật toán di truyền mới cho bài toán lập lịch job shop (Jobshop Scheduling Problem - JSP). Thuật toán này có một số đổi mới sau đây: Một lịch biểu được mã hóa bởi các số tự nhiện; các luật ưu tiên của Giffler và Thompson được dùng để tạo ra các lịch biểu tích cực; phép đột biến được thực hiện trên các cá thể tiềm năng và kết hợp với kĩ thuật tìm kiếm lân cận; phép trao đổi chéo mới kết hợp trao đổi chéo đồng nhất với thuật toán GAs và được thực hiện trên 3 cá thể cha.

Bài báo nghiên cứu về bài toán thời khóa biểu: “Một giải pháp tiến hóa cho bài toán thời khóa biểu” . Bài báo giới thiệu một phương án tiến hóa cho bài toán lập thời khóa biểu tại các trường phổ thông.

Bài viết về bài toán đóng thùng: “Về thuật toán di truyền lai giải bài toán đóng thùng”, trình bày một số kết quả liên quan đến việc xây dựng, cài đặt và thử nghiệm thuật toán di truyền kết hợp với các thuật toán tìm kiếm trực tuyến oristic.

Bài viết khác về bài toán phân hoạch: “Hướng tiếp cận mới trong việc tách từ để phân loại văn bản tiếng Việt sử dụng giải thuật di truyền và thống kê trên Internet” . Giải thuật di truyền áp dụng cho bài toán tách từ tiếng Việt được tóm tắt như sau: Xét văn bản t gồm n tiếng t=s1s2…sn. Mục tiêu của GAs là xác định những cách tách hợp lý nhất văn bản t thành m đọan t=w1w2…wm với wk=si…sj (1 ≤ k≤ m, 1≤ i, j≤ n) có thể là từ đơn hay từ phức.

Ứng dụng của GAs trong học tham số cho mạng noron: “Điều khiển Robot với mạng Noron RBF có hệ số học được tối ưu bằng giải thuật di truyền” . Đối với các hệ thống điều khiển có tính phi tuyến cao và độ bất định lớn, các tham số của hệ thống như độ ma sát, sự thay đổi của tải trọng ... luôn thay đổi trong trong quá trình điều khiển do đó yêu cầu rất quan trọng trong quá trình học của các bộ điều khiển nơrơn là phải đảm bảo quá trình học online. Trong các nghiên cứu gần đây về quá trình học của mạng nơron dù bằng phương pháp hàm bán kính cơ sở hay bằng phương pháp EBP thì hệ số học h của mạng nơron vẫn được chọn theo kinh nghiệm của người lập trình. Trong bài báo này tác giả đề cập đến việc sử dụng công cụ GAs để tối ưu hệ số học của mạng nơron RBF thay vì lựa chọn bằng kinh nghiệm.

# Chương 2: Bài toán lập lịch

## 2.1. Tìm hiểu chung

Lập lịch có thể được định nghĩa là một bài toán tìm kiếm chuỗi tối ưu để thực hiện một tập các hoạt động chịu tác động của một tập các ràng buộc cần phải được thỏa mãn. Người lập lịch thường cố gắng thử đến mức tối đa sự sử dụng các cá thể, máy móc và tối thiểu thời gian đòi hỏi để hoàn thành toàn bộ quá trình nhằm sắp xếp lịch.Vì thế bài toán lập lịch là một vấn đề rất khó để giải quyết. Hiện nay có nhiều khả năng để phát triển các kỹ thuật hiện tại để giải quyết bài toán này. Những kỹ thuật đó bao gồm: các tiếp cận Trí tuệ nhân tạo như hệ thống tri thức cơ sở (knowledge-based systems), bài toán thỏa mãn ràng buộc, hệ chuyên gia, mạng Nơron và các tiếp cận của các Nghiên cứu hoạt động: lập trình tính toán, lập trình động, tìm kiếm nhánh và đường biên, kỹ thuật mô phỏng, tìm kiếm Tabu và phương pháp nút cổ chai

Bài toán thời khóa biểu là một bài toán có sự kết hợp giữa nhiều loại ràng buộc khác nhau, các ràng buộc đều phải chặt chẽ và thống nhất. Có nhiều phiên bản của bài toán thời khóa biểu, một trong những bài toán này có thể được mô tả như sau:

Có một danh sách các giáo viên, một danh sách các khoảng thời gian, một danh sách các lớp, danh sách các phòng học. Bài toán cần tìm thời khóa biểu tối ưu (giáo viên – thời gian – lớp - phòng học); hàm mục tiêu phải thỏa những mục tiêu này (các ràng buộc mềm) gồm: Có một số giờ được xác định trước cho mỗi giáo viên và mỗi lớp; Chỉ một giáo viên trong một lớp vào một giờ nhất định; Một giáo viên không thể dạy hai lớp cùng lúc; một phòng học không thể có hai lớp học hai thầy cô khác nhau;  Đối với mỗi lớp được xếp thời khóa biểu vào một khoảng thời gian, phải có một giáo viên…

## 2.2. Mô tả bài toán lập lịch tổng quát

**Vấn đề đặt ra**

Bài toán cần sắp xếp thời khóa biểu cho một trường đại học với nhiều cơ sở khác nhau, cần có sự sắp xếp lịch học cho các lớp tại các phòng ở mỗi địa điểm sao cho phù hợp nhất.

**Dữ liệu Bài toán:**

* Danh sách các bộ môn
* Danh sách các giáo viên
* Danh sách các lớp học
* Danh sách các phòng học
* Danh sách giờ học

**Các bước cơ bản trong bài toán Lập Lịch TKB**

Bước 1: Khởi tạo dữ liệu thời khóa biểu mới

Bước 2: Nhập, điều chỉnh dữ liệu gốc thời khóa biểu

Bước 3: Nhập, sửa, điều chỉnh các ràng buộc của thời khóa biểu

Bước 4: Chuẩn bị xếp thời khóa biểu

Bước 5: Xếp tự động Thời khóa biểu

Bước 6: Điều chỉnh TKB

Bước 7: Hoàn Thiện TKB

**Từ các dữ liệu trên, ta có thể mô tả tổng quát về bài toán lập lịch**

Cho một tập n công việc {Ji}1 ≤ i ≤ n, mỗi công việc bao gồm m thao tác được xử lý ở trên một tập m máy {Mj}1 ≤ j ≤ m và thỏa mãn các ràng buộc sau đây:

 1. Mỗi công việc phải được xử lý ở trên mỗi máy theo một trình tự cho trước của các thao tác. Trình tự thực hiện các thao tác của mỗi công việc lần lượt trên các máy được gọi là tuần tự công nghệ.

 2. Tại một thời điểm mỗi máy chỉ có thể xử lý nhiều nhất là một công việc.

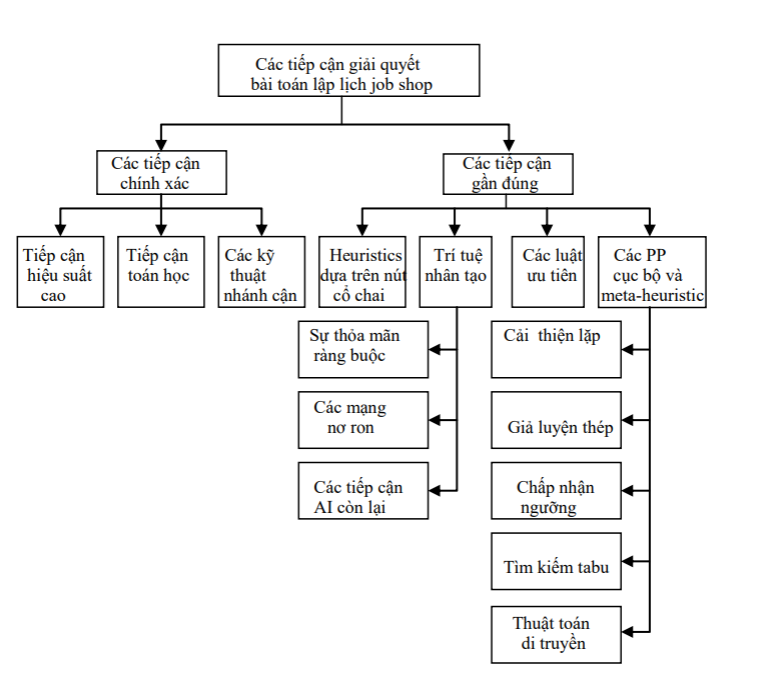
3. Mỗi máy Mj tùy ý đều có khả năng xử lý một công việc Ji nào đó, phần công việc Ji được xử lý trên máy Mj được gọi là thao tác Oij.

 4. Mỗi thao tác Oij phải được xử lý một cách liên tục ở trên máy Mj (không bị ngắt khi đang xử lý).

5. Thời gian bắt đầu xử lý và thời gian hoàn thành việc xử lý thao tác Oij được ký hiệu lần lượt là sij và cij. Thời gian xử lý thao tác Oij được ký hiệu là pij.

6. Thời gian hoàn thành việc xử lý tất cả các công việc được gọi là makespan và được ký hiệu là Cmax.

## **2.3. Những giải pháp thuật toán đã được đề ra cho bài toán Lập Lịch Thời** Khóa Biểu

****

-> Tuy nhiên luôn có những tồn đọng trong các giải pháp được đề xuất này.

Những vấn đề chưa giải quyết được:

* Việc đánh giá kết quả của các giải pháp khác nhau đòi hỏi phải chính xác, đặc biệt các kết quả có liên quan tới heuristic. Các thiết kế thử nghiệm hiện nay còn có nhiều bất cập, các thiết kế thử nghiệm này cần phải nghiêm ngặt hơn, các bài toán test chuẩn cần phải nhiều hơn và đa dạng hơn.
* Cần phải chứng tỏ thuộc tính hội tụ tới tối ưu toàn cục của các tiếp cận gần đúng được đề xuất mới cho bài toán trên cơ sở lý thuyết.
* Chưa có một phương pháp hình thức được được đề xuất cho việc kết hợp hiệu quả các kỹ thuật tìm kiếm với nhau. Các nghiên cứu nên đưa ra cách thức kết hợp như thế nào để có một tiếp cận mới mạnh hơn.
* Cuối cùng là sự phức tạp của các bài toán lập lịch cỡ lớn với không gian lời giải quá lớn của nó. Do vậy, nên tập trung vào các kỹ thuật tìm kiếm song song trên nhiều vùng lân cận của không gian lời giải.

Từ những bất cập trên thì chúng ta nên có giải pháp nào tốt hơn để giải quyết triệt để cho bài toán này? Chúng ta cùng qua Chương 2 để tìm hiểu rõ hơn về các cài đặt thuật toán và áp dụng vào bài toán như thế nào.

# Chương 3. Cài đặt thuật toán di truyền và bài toán lập lịch thời khóa biểu

## 3.1. Chọn mô hình cá thể

Lịch của một môn học sẽ gồm tên môn học, tên bộ môn phụ trách, mã giảng viên dạy môn đó và thời gian học của khóa học đó. Như vậy đối với một bộ môn sẽ phụ trách nhiều môn học khác nhau. Đối với mỗi môn học, việc làm nhiễm sắc thể là phù hợp với tính không ổn định của nó, với số lượng các môn phụ thuộc từng lớp học, cũng giống như số lượng nhiễm sắc thể trong cá thể, có chiều dài không nhất thiết phải cố định hay bằng nhau. Ngoài ra tính phức tạp của môn học về số tiết phải học luôn bị thay đổi, trong khi giá trị các giờ học thì ngược lại, có thể xác định một cách rõ ràng và nhanh chóng.

Mô tả cá thể trong lịch:

|  |  |
| --- | --- |
| Tên môn học | Các thuộc tính |
| Môn học 1 |  |
| Môn học 2 |  |
| …... |  |
| Môn học n |  |

Thay vì chọn ngẫu nhiên môn học vào các tiết học như đã trình bày, chúng ta sẽ làm ngược lại: chọn ngẫu nhiên tiết học theo môn, vì chúng ta đã chọn môn học làm đơn vị trong cá thể ( theo mô hình trên ). Có nghĩa là, với một cá thể của mô hình xếp lịch lớp, ở bất kỳ thời điểm nào, khi ta đặt nhiễm sắc thể đầu tiên như là môn thứ nhất, nhiễm sắc thể kế tiếp sẽ là môn thứ hai, và tiếp tục cho các nhiễm sắc thể còn lại ... thì sau này, lúc nào cũng theo thứ tự ấy mà lấy thông tin ra, sẽ không có gì thay đổi ( ngoại trừ giá trị tiết học, nếu như sau này có xảy ra lai ghép hay đột biến ). Trong trường hợp một môn được học nhiều lần trong tuần, do có nhiều chứng chỉ / học phần, nên sẽ gây khó khăn cho việc xếp chúng vào trong cá thể. Cách giải quyết vấn đề này, chỉ cần đưa chúng vào cá thể với nhiễm sắc thể tương ứng, chẳng khác gì một môn học bình thường khác. Lúc đọc thông tin, chúng ta nên chú ý một chút thế thôi.

Ví dụ: Có danh sách các môn học và có số lần học như sau:

* Môn học a có 1 buổi/tuần
* Môn học b có 2 buổi/tuần
* Môn học c có 1 buổi/tuần

Ta sẽ chia các nhiễm sắc thể như sau:

|  |  |
| --- | --- |
| Tên môn học | Các thuộc tính |
| Môn học a |  |
| Môn học b |  |
| Môn học b |  |
| Môn học c |  |

Trong thuộc tính của mỗi nhiễm sắc thể sẽ chứa thời gian học, mã giảng viên và bộ môn phụ trách. Do vậy, khi tạo cá thể, chúng ta vẫn phải đảm bảo sự đúng đắn về tính chất trong lịch học: Số lượng tốt đa sinh viên đi học khóa học đó phải nhỏ hơn sức chứa của phòng học, thời gian học của các môn không bị trùng nhau khi chúng cùng một phòng và cùng một giảng viên,...

## 3.2. Tạo quần thể ban đầu

Trước khi tạo quần thể ban đầu trong phần này, chúng ta phải chuẩn bị sẵn về dữ liệu cho quá trình thực thi, từ lúc khởi tạo đến khi cho ra kết quả, bao gồm đầy đủ thông tin của một lớp đang được chọn. Tất cả như sau :

* Các ràng buộc trong lịch của một môn học, giảng viên được phân công dạy.
* Các môn học và mã của cùng giảng viên phụ trách.
* Các phương thức đọc và ghi nhận nhiễm sắc thể.

Sau khi tạo ngẫu nhiên quần thể ban đầu, ta cần đánh giá quần thể, kiểm tra xem độ thích nghi của quần thể đó có tốt hay không.

## 3.3. Độ thích nghi – Chọn cá thể

Giai đoạn hội tụ cá thể trong quần thể, trên cơ bản việc đánh giá tùy thuộc vào một số yếu tố từ đó xác định các ràng buộc.Bằng cách đánh giá và xét dựa trên các ràng buộc, ứng với một ràng buộc sẽ được gán bởi một giá trị thích nghi tương ứng, khi cá thể đi qua, các ràng buộc được lắp đặt vào, và sẽ cho ra giá trị thích nghi cụ thể cho cá thể đó. Để tính toán độ thích nghi, ta xét một số ràng buộc sau:

* Đầu tiên là về giảng viên: Khi phân công giảng dạy, phải chắc chắn rằng giảng viên đó sẽ trống vào giờ đó, môn đó và buổi đó. Trước tiên là kiểm tra thời gian học của các môn học có trùng nhau không sau đó kiểm tra phòng học có bị chồng chéo lẫn nhau hay giảng viên có bị trùng khi một thời gian mà một giảng viên dạy hai lớp (Chắc chắn là thực tế có thể sảy ra trường hợp này tùy vào yêu cầu).
* Trong khi chọn lai ghép, có thể dẫn tới tính đột biến gây ra sai lệch về mặt ràng buộc. Cho nên ta phải kiểm tra đầy đủ những ràng buộc mà yêu cầu đặt ra.

Như đã giới thiệu ở phần 1, có rất nhiều cách để chọn một cá thể tốt. Ở đây chọn theo độ thích nghi của cá thể tốt nhất cũng là một phương án tốt. Chúng ta sẽ xếp theo giá trị giảm dần của độ thích nghi.

## 3.4. Thuật toán lai ghép – đột biến

Về thuật toán lai ghép, ta dùng lai ghép theo đoạn: ý là sẽ lấy ngẫu nhiên một đoạn nhiễm sắc thể làm nhiễm sắc thể cha và còn lại sẽ làm nhiễm sắc thể mẹ.

Đối với đột biến: Chỉ cần hoán vị hai nhiễm sắc thể một cách ngẫu nhiên trong cá thể. Ta có thể sửa thông số xác suất về đột biến để phù hợp với yêu cầu.

## 3.5. Chọn điểm dừng thuật toán

Như đã nêu ở trên, thuật toán di truyền sẽ tính toán độ thích nghi sau mỗi lần lai ghép để chọn con tốt nhất. Do vậy độ thích nghi cũng có thể được xác định như là một điểm dừng của thuật toán. Một số bài toán chọn điểm dừng theo số thế hệ, hoặc dựa trên tính tương đối của kết quả, nhưng với bài toán này cần có một kết thúc tuyệt đối tốt nhất, mặt dù số thế hệ vẫn phải được chọn trước ngay từ đầu.Vì tính chất yêu cầu trong bài toán này là không bị sai lệch.

Do vậy nếu trong quá trình thực thi qua các giai đoạn, chỉ cần một kết quả không đạt đến điểm dừng, xem như bài toán sẽ không có kết quả.

# Chương 4. Thiết kế hệ thống lập lịch thời khóa biểu

## 4.1. Thiết kế cơ sở dữ liệu rang buộc của bài toán

Khi tạo một lịch cho lớp học chúng ta cần xem xét đến những dữ liệu trong đó. Ở đây nhóm em thực hiện việc lưu dữ liệu dưới dạng file excel.

|  |  |
| --- | --- |
|  |  |
|  |  |

Đầu vào của thuật toán bao gồm: Danh sách các bộ môn, danh sách các môn học, danh sách giảng viên, phòng học và thời gian học. Các yêu cầu này sẽ được chia thành từng nhóm và tùy theo tầm quan trọng của chúng. Các yêu cầu bắt buộc (nếu vi phạm một trong các ràng buộc này, thì lịch đó sẽ không được chấp nhận.

* Số lượng sinh viên tốt đa của môn học lớn hơn sức chứa của phòng học đó.
* Không có giảng viên nào có thể dạy nhiều lớp học cùng một lúc.
* Không có môn học nào chứa nhiều lớp cùng một lúc.

Việc lấy dữ liệu từ file excel ứng với mỗi hàng của excel là một đối tượng được tạo ra nhằm làm đầu vào cho thuật toán.

## 4.2. Các thành phần đối tượng của bài toán

Giảng viên: Phần giảng viên có ID( chỉ danh) và tên của giảng viên.

Phòng học: Phần phòng học chứa ID và tên của phòng học cũng như sức chứa của lớp học đó có thể đạt được.

Bộ môn: Phần bộ môn chứa mã bộ môn, tên bộ môn và danh sách các môn học mà bộ môn đó phụ trách.

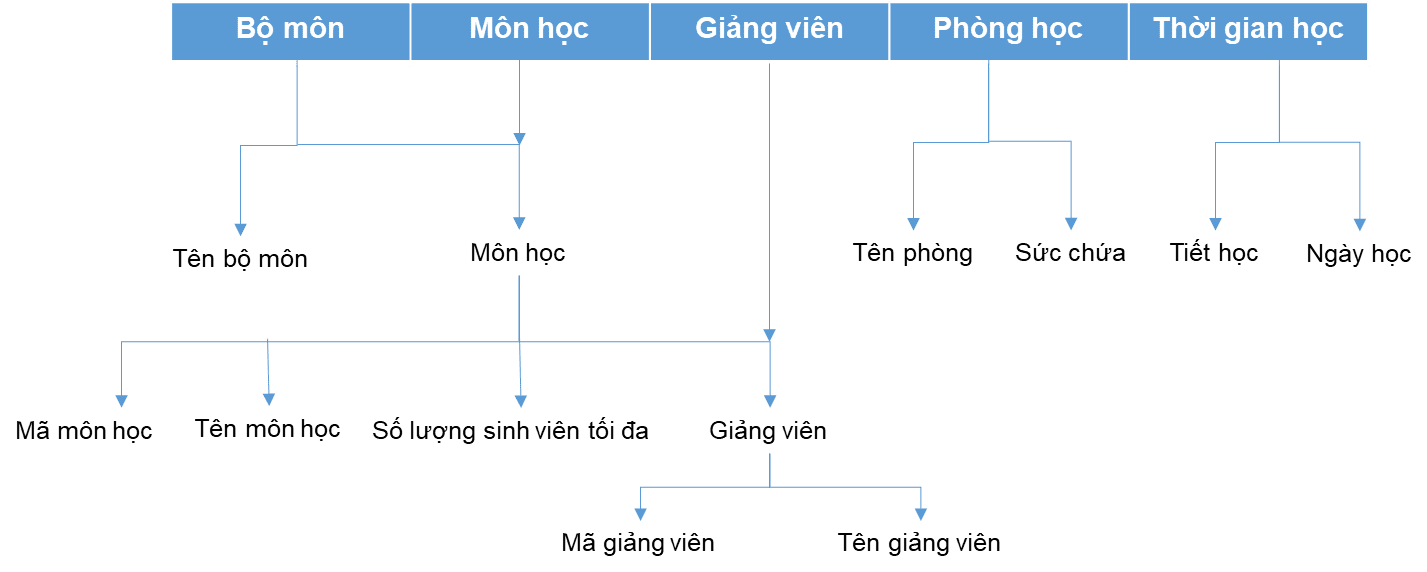
Thời gian học: Được chia thành hai thành phần:

* Thành phần thứ nhất là: Ca học
* Thành phần thứ nhất là: Ngày học (ở đây một ngày sẽ gồm 4 ca)

Môn học: Đối với môn học sẽ bao gồm mã môn học, mã giảng viên phụ trách và số lượng sinh viên tối đa có thể học môn đó.

## 4.3. Thiết kế nhiễm sắc thể

Sau khi nhận được đầu vào là dữ liệu, thông tin của môn học và phòng học. Chúng ta tạo ra một đối tượng Schedule để gom nhóm lưu trữ toàn bộ những thông tin trên.



*Biểu diễn nhiễm sắc thể (lưu trong thuộc tính \_classes).*

Bên cạnh đó mỗi nhiễm sắc được lưu trữ những giá trị phù hợp để xác định độ thích nghi dựa trên những conflict được xác định bởi những điều kiện của đề bài. Giá trị thích nghi được lưu tại biến **\_fitness** và ban đầu thì độ thích nghi là -1.0. Độ thích nghi tốt nhất là 1.0.

## 4.4. Các tham số của hệ thống và thao tác lai ghép

Số lượng nhiễm sắc thể trong một quần thể.

POPULATION\_SIZE = 9

Số lượng quần thể được dữ lại khi tiến hành lai ghép:

NUMBER\_OF\_ELITE\_SCHEDULES = 1

Số lượng lai ghép trong các lớp cha mẹ:

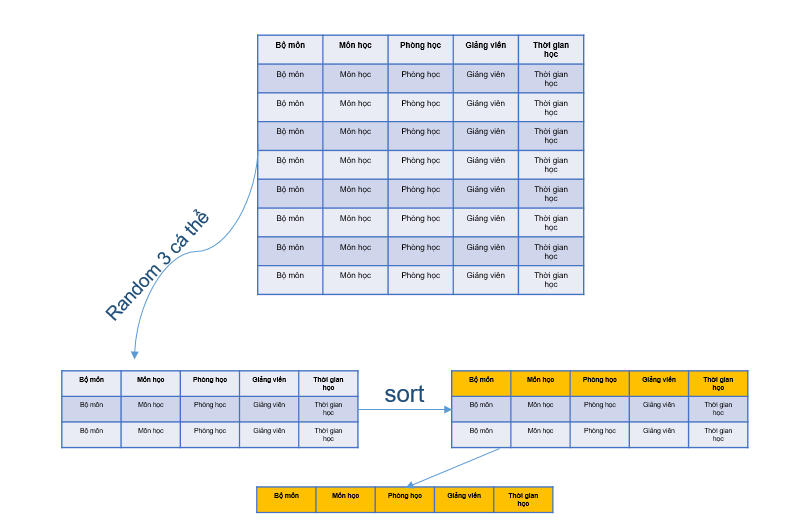
TOURNAMENT\_SELECTION\_SIZE = 3

Tỉ lệ đột biến sẽ xảy ra:

MUTATION\_RATE = 0.1

### 4.4.1. Phép chọn cá thẻ vào lớp cha-mẹ

Sau khi tạo các quần thể xong, sẽ chọn n cá thể từ quần thể tính độ thích nghi cho từng cá thể sử dụng phương pháp tournament để chọn cá thể có độ thích nghi tốt nhất trong n cá thể được chọn và đưa vào lớp cha mẹ để tiến hành lai ghép.

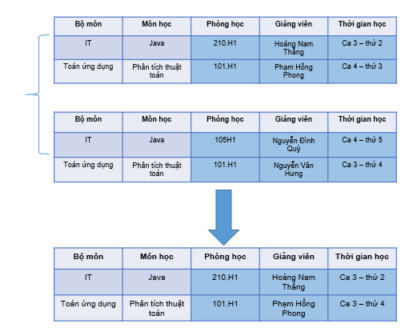


### 4.4.1. Phép lai ghép

Như đã nói ở trên phép lai ghép diễn ra bằng cách ghép một hành nhiều đoạn gen từ hai nhiễm sắc thể cha và mẹ để hình thành nhiễm sắc thể mới mang đặc tính của cả cha lẫn mẹ.

Đầu tiên tiến hành chọn hai lớp cha và mẹ được chọn ngẫu nhiên từ 3 nhiễm sắc thể trong quần thể và lấy hai nhiễm sắc thể có độ tương thích cao nhất. Cho chúng lai ghép với nhau.

Ví dụ:



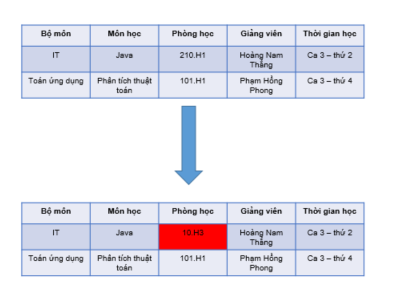
Ý tưởng của phép lai ghép sẽ chọn một xác suất xem nên nghiêm về lớp cha hay lớp mẹ nhiều hơn - ở đây nhóm em nhận thấy rằng hai cá thể cha và mẹ thì có tính chất giống nhau nên sẽ để xác suất là 50% lớp cha và 50% lớp mẹ. Sau khi lai ghép sẽ trả ra một nhiễm sắc thể mới được tạo ra từ hai cá thể cha-mẹ.

Quá trình lai ghép sẽ giữ lại một nhiễm sắc thể và tiến hành chọn những cặp cha-mẹ trong những nhiễm sắc thể còn lại. Sau đó thay đổi luân phiên các nhiễm sắc thể cha-mẹ thành những nhiễm sắc thể mới và các hình thức thao tác lai ghép.

### 4.4.2. Phép đột biến

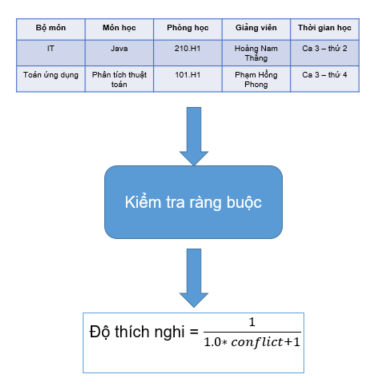
Cũng giống như phép lai ghép, phép đột biến biểu diễn ra bằng cách khi một hoặc một số tình trạng của con không được thừa hưởng từ hai chuỗi nhiễm sắc thể cha mẹ. Phép đột biến xảy ra với xác suất thấp hơn rất nhiều so với xác suất xảy ra phép lai ghép.

Trong hệ thống này thao tác đột biến ban đầu kiểm tra xác suất đột biến với một số bất kì



Sau đó thao tác đột biến tạo ra một lớp ngẫu nhiên và di chuyển nhiễm sắc thể đến một slot cũng được lựa chọn ngẫu nhiên khác. Số của lớp học đó sẽ được di chuyển vào một thao tác đơn lẻ được xác định bởi kích thước đột biến trong các tham số của nhiễm sắc thể

## 4.5. Độ thích nghi của hệ thống



Đây là phần giải quyết các yêu cầu đưa ra cho bài toán, chủ yếu vẫn xem xét trên các thành phần ràng buộc. Tương ứng với mỗi loại ràng buộc chúng ta sẽ xác định cho nó số conflict. Khi đi qua các ràng buộc sẽ cho ra giá trị thích nghi cụ thể cho cá thể đó, kết thúc công việc tính độ độ thích nghi.

Một số ràng buộc trong bài toán:

* Số lượng sinh viên trong môn học không được vượt quá số sức chứa của phòng học.
* Khi cùng một thời gian 2 lớp khác nhau sẽ không vào cùng một phòng học và không cùng giảng viên.

# Chương 5. Các thực nghiệm của hệ thống

Thực nghiệm được tiến hành trên những trường hợp dữ liệu đầu vào là khác nhau từ đó tính toán độ thích nghi sau mỗi lần tiến hóa.

*Cấu hình máy thực hiện thực nghiệm:*

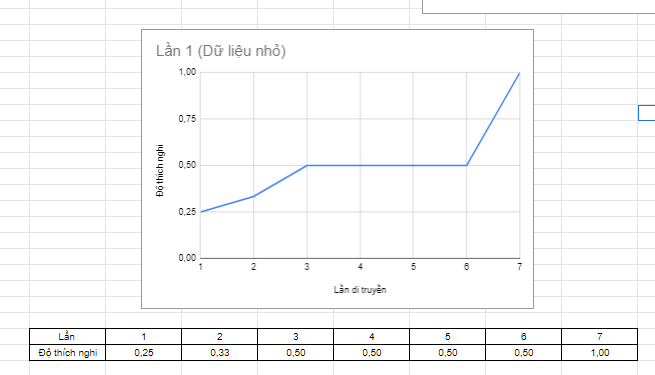
* CPU: Intel(R) Core(TM) i7 8565U CPU @ 1.80GHz 1.99 GHz.
* RAM: 8 GB DDR3L 1800MHz bus.
* VGA: [NVIDIA® GeForce® GT 720M](https://www.laptopvip.vn/nvidia-geforce-gt-720m-vi.html)
* Hệ điều hành Windows 10 Pro.
* Compiler: Visual Studio code.

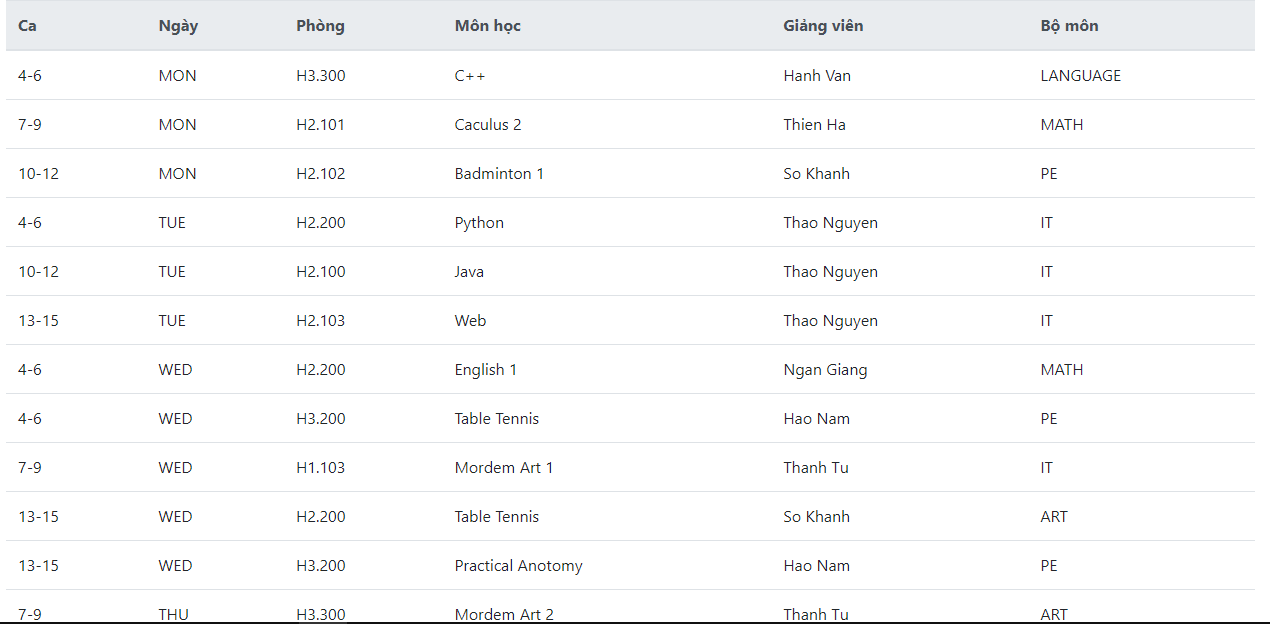
*Ngôn ngữ cài đặt:*

* Python 3.10 (latest version).

**Lần 1.**

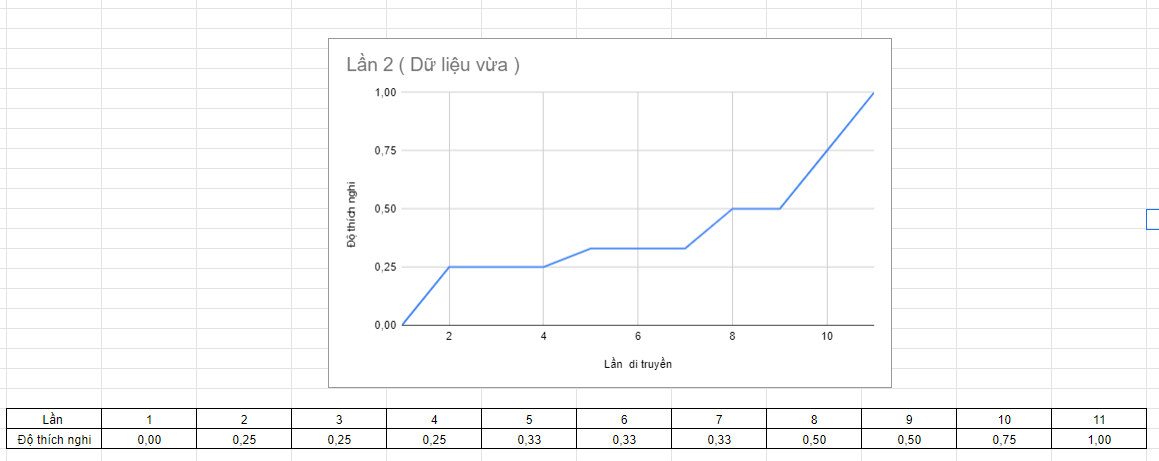
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dữ liệu | Số môn học | Số giảng viên | Số phòng học | Cỡ quần thể | Tỉ lệ đột biến |
| Số lượng | 16 | 12 | 16 | 10 | 0.1 |





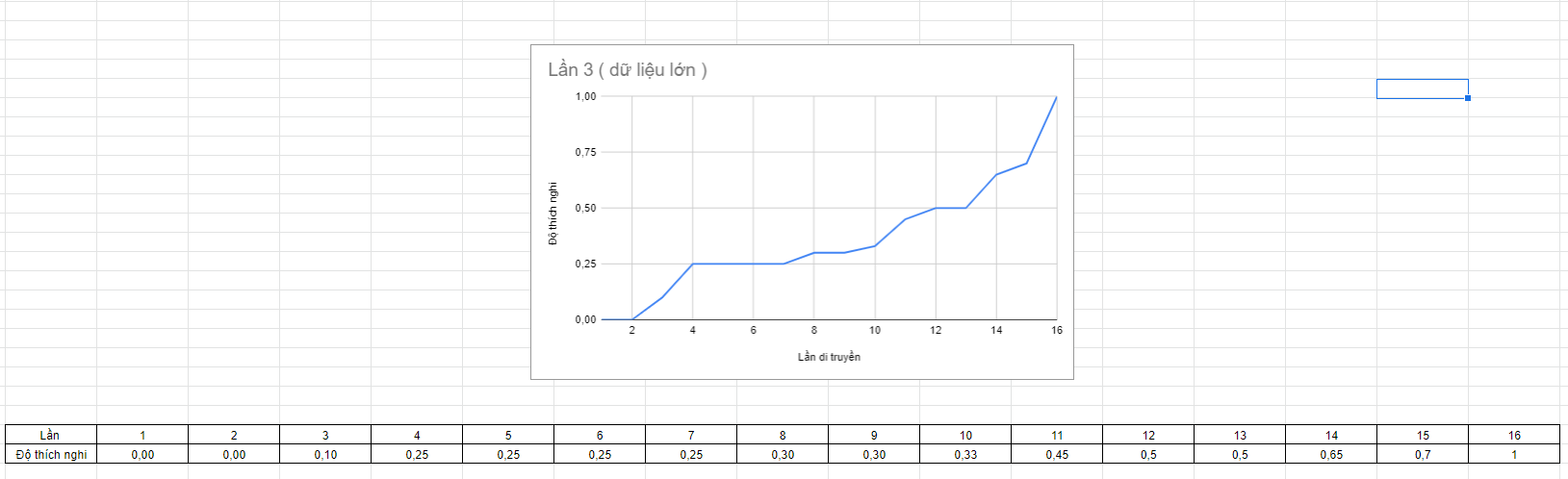
**Lần 2:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dữ liệu | Số môn học | Số giảng viên | Số phòng học | Cỡ quần thể | Tỉ lệ đột biến |
| Số lượng | 25 | 16 | 18 | 11 | 0.1 |



**Lần 3:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dữ liệu | Số môn học | Số giảng viên | Số phòng học | Cỡ quần thể | Tỉ lệ đột biến |
| Số lượng | 35 | 16 | 20 | 11 | 0.1 |



# Chương 6. Kết luận và đề xuất hướng phát triển.

Giải thuật di truyền (GAs) trong lĩnh vực tin học là một trong những giải thuật thú vị, bởi vì nó mô phỏng qui luật đấu tranh sinh tồn của tự nhiên và cũng là một giải thuật vô cùng hiệu quả đối với các loại bài toán tối ưu. Trong bài viết này, chúng tôi trình bày những đặc điểm cơ bản nhất của GAs, nêu một vài ứng dụng và một số công trình nghiên cứu về GAs đã được công bố ở trong nước.

Bài toán tối ưu là bài toán tìm kiếm giải pháp (tốt nhất) trong không gian (vô cùng lớn) các giải pháp. Khi không gian tìm kiếm nhỏ, các phương pháp cổ điển cũng đủ thích hợp; nhưng khi không gian lớn hơn thì cần phải dùng đến những kĩ thuật đặc biệt, GAs là một trong những kĩ thuật đó.

GAs là hệ thống đầu tiên dựa trên mô phỏng sự sống. Thông qua chọn lọc, lai ghép, đột biến, GAs sẽ hội tụ qua nhiều thế hệ theo hướng tối ưu toàn cục. GAs hi vọng hướng tới một giải pháp tối ưu hơn bằng cách kết hợp thông tin tốt ẩn trong tập hợp các giải pháp, để tạo ra giải pháp mới với những thông tin tốt thừa hưởng từ cả cha và mẹ. Khả năng của thuật toán vẫn đang còn được khám phá và khai thác, một số lượng lớn những nghiên cứu thành công về lí thuyết cũng như ứng dụng trong nhiều lĩnh vực thế giới thực đã chứng minh rằng GAs là một kĩ thuật tối ưu mạnh mẽ và thiết thực.

## 6.1. Kết quả đạt được

Tìm hiểu về các ứng dụng của thuật toán GAs.

Một số công trình đã được nghiên cứu.

Áp dụng được giải thuật di truyền để giải quyết bài toán sắp thời khoá biểu. Xây dựng thành công chương trình demo sắp xếp thời khoá biểu.

## 6.2. Hạn chế - Hướng phát triển trong tương lai.

**Hạn chế:**

+ Do giải thuật di truyền mang tính chất ngẫu nhiên nên đôi khi kết quả đạt được không phải là 100%.

+ Giải thuật Di Truyền có thể giải quyết bài toán tối ưu bất kỳ (cực tiểu hóa hàm mục tiêu) với n biến vào. Tuy nhiên, với số lượng biến vào khá nhiều, các giá trị hàm mục tiêu đạt được thường không gần với kết quả tối ưu thực sự.

Để khắc phục vấn đề này, có thể tăng số lượng vòng lặp, hy vọng lần sinh sản muộn sẽ hình thành những con cháu với độ thích nghi cao ứng với các giá trị hàm mục tiêu gần kết quả tối ƣu thực sự nhất.

**Hướng phát triển trong tương lai:**

+ Sắp thời khóa biểu thực hành theo nhiều mức độ ưu tiên hơn( ưu tiên cho giảng viên….).

+ Hoàn thiện một số các chức năng hiệu chỉnh để người dùng có thể linh động hơn trong quá trình hiệu chỉnh.

# Tài liệu tham khảo.

* Lập trình tiến hoá\_Ts. Nguyễn Đình Thúc.
* Giới thiệu giải thuật Di truyền và Tính toán Tiến hóa \_PGS.TS Randy Ribler khoa tin trường đại học Lynchburg,VA,USA.
* Paper: Genetic Algorithm Proceedings published by International Journal of Computer Applications