

**TRƯỜNG ĐẠI HỌC XÂY DỰNG
KHOA CÔNG NGHỆ THÔNG TIN**

-----o0o-----



BÀI NGHIÊN CỨU

ĐỀ TÀI: Thuật toán di truyền và bài toán lập lịch

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
Mở đầu	5
Lý do chọn đề tài.....	5
Mục tiêu của bài tập nghiên cứu	6
Ý nghĩa khoa học và ý nghĩa thực tiễn	6
Chương 1. Tổng quan về thuật toán di truyền và bài toán lập lịch.....	7
1.1. Thuật toán di truyền	7
1.1.1. Giới thiệu về thuật toán.....	7
1.1.2. Cấu trúc của thuật toán di truyền	8
1.1.3. Mã giả của thuật toán	10
1.1.4. Các tham số, công thức của giải thuật	10
1.1.5. Các thành phần của giải thuật di truyền.....	11
1.1. Bài toán lập lịch.....	13
1.2.1. Tìm hiểu chung	13
1.2.2. Mô tả bài toán lập lịch tổng quát	14
1.2.3. Những giải pháp thuật toán đã được đề ra cho bài toán Lập Lịch Thời Khóa Biểu	15
Chương 2. Cài đặt thuật toán di truyền và bài toán lập lịch thời khóa biểu	16
2.1. Chọn mô hình cá thể	16
2.2. Tạo quần thể ban đầu	17
2.3. Độ thích nghi – Chọn cá thể	18
2.4. Thuật toán lai ghép – đột biến	18
2.5. Chọn điểm dừng thuật toán.....	19
Chương 3. Thiết kế hệ thống lập lịch thời khóa biểu	20
3.1. Thiết kế cơ sở dữ liệu ràng buộc của bài toán	20
3.2. Các thành phần đối tượng của bài toán.....	21
3.3. Thiết kế nhiệm sắc thể	21
3.4. Các tham số của hệ thống và thao tác lai ghép	22

3.4.1. Phép chọn cá thể vào lớp cha-mẹ	22
3.4.1. Phép lai ghép.....	23
3.4.2. Phép đột biến.....	24
3.5. Độ thích nghi của hệ thống	25
Chương 4. Các thực nghiệm của hệ thống.....	26
Chương 5. Kết luận và đề xuất hướng phát triển.....	28
5.1. Kết quả đạt được	28
5.2. Hạn chế - Hướng phát triển trong tương lai.....	28
Chương 6. Tài liệu tham khảo.	29

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 tự 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/vết cặn (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.
2. 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 và bài toán lập lịch.

Chương 2: Cài đặt thuật toán di truyền vào bài toán xếp thời khóa biểu.

Chương 3: Thiết kế hệ thống lập lịch thời khóa biểu.

Chương 4: Các thực nghiệm thực tế.

Chương 5: Kết luận và đề xuất hướng phát triển.

Chương 1. Tổng quan về thuật toán di truyền và bài toán lập lịch

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.

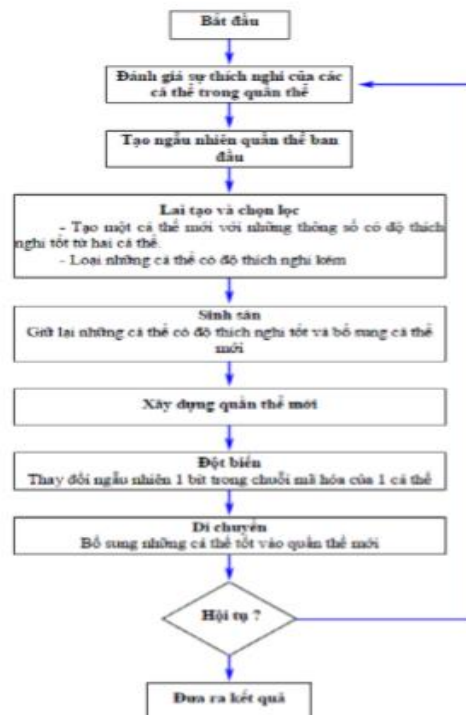
Thuật toán di truyền (Genetic Algorithms) là thuật toán được mô phỏng theo quá trình thích nghi tiến hóa, chọn lọc tự nhiên, là kỹ thuật chung giúp giải quyết vấn đề bài toán bằng cách mô phỏng sự tiến hóa của con người hay của sinh vật nói chung (dựa trên thuyết tiến hóa của Darwin hay luật di truyền 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 hóa: 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 tùy 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:

$t = t + 1$

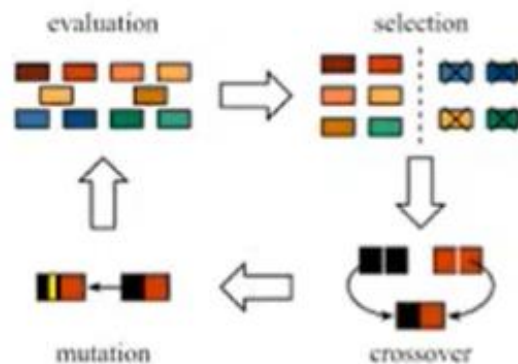
Lựa chọn $P(t)$

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 được 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 được 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.

$$\text{Độ thích nghi} = \frac{1}{1.0 * \text{conflict} + 1}$$

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

1.1. Bài toán lập lịch

1.2.1. 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 Noron 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...

1.2.2. 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

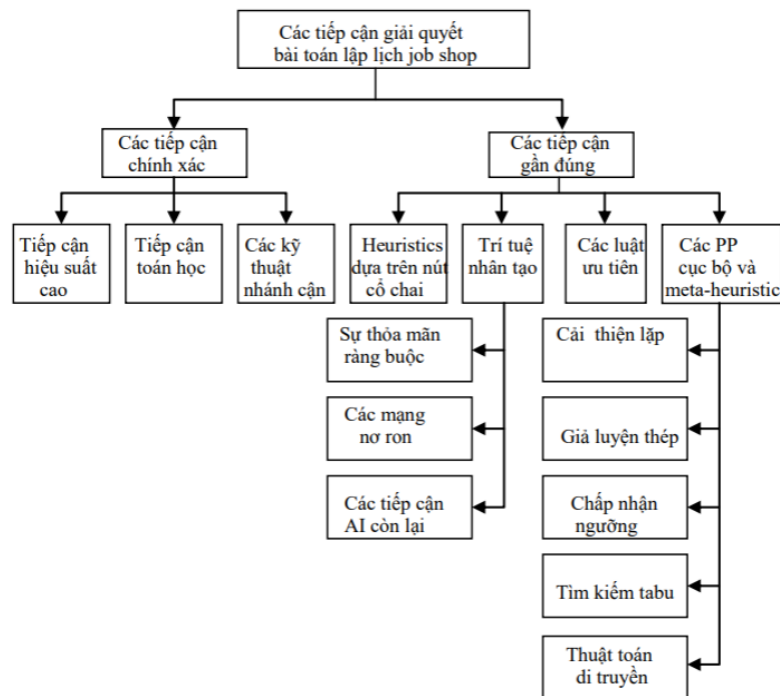
Từ các dữ liệu trên, ta có thể mô tả tổng quát về bài toán lập lịch

Cho một tập n công việc $\{J_i\} 1 \leq i \leq n$, mỗi công việc bao gồm m thao tác được xử lý ở trên một tập m máy $\{M_j\} 1 \leq j \leq m$ 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 M_j tùy ý đều có khả năng xử lý một công việc J_i nào đó, phân công việc J_i được xử lý trên máy M_j được gọi là thao tác O_{ij} .
4. Mỗi thao tác O_{ij} phải được xử lý một cách liên tục ở trên máy M_j (không bị ngắt khi đang xử lý).

5. Thời gian bắt đầu xử lý và thời gian hoàn thành việc xử lý thao tác O_{ij} được ký hiệu lần lượt là s_{ij} và c_{ij} . Thời gian xử lý thao tác O_{ij} được ký hiệu là p_{ij} .
6. Thời gian hoàn thành việc xử lý tất cả các công việc được gọi là makespan và được ký hiệu là C_{max} .

1.2.3. 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 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 2. Cài đặt thuật toán di truyền và bài toán lập lịch thời khóa biểu

2.1. 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 nhiệm sác 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 nhiệm sác 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 nhiệm sắc thể đầu tiên như là môn thứ nhất, nhiệm sắc thể kế tiếp sẽ là môn thứ hai, và tiếp tục cho các nhiệm 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 nhiệm sắc thể tương ứng, chẳng khác gì một môn học bình thường khác. Lúc đọc thông tin, chúng ta nên chú ý một chút thể 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 nhiệm 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 nhiệm 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 tối 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,...

2.2. 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 nhiệm 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.

2.3. 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 xảy 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.

2.4. 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.

2.5. 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 3. Thiết kế hệ thống lập lịch thời khóa biểu

3.1. Thiết kế cơ sở dữ liệu ràng 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.

	A	B	C	D	E	F
1	id	name	instructor	max_number_of_students		
2	0	Caculus 1	1,2	30		
3	1	Caculus 2	2	30		
4	2	English 1	3	45		
5	3	English 2	4,5	45		
6	4	Chinese	3,5	30		
7	5	C++	6,8	60		
8	6	Web	7,6	30		
9	7	Python	7	35		
10	8	Java	7	40		
11	9	Mordem A	9,10	35		
12	10	Mordem A	10	35		
13	11	Practical A	11	55		
14	12	Table Tenn	11,12,12	45		
15	13	Baseball	11,12	40		
16	14	Badminton	11,12	40		

	A	B
1	name	course_id
2	MATH	1,2
3	LANGUAGE	3,4,5
4	IT	6,7,8,9
5	ART	10,11,12
6	PE	11,12,14

	A	B	C
1	id	professor	
2	0	Hang Nga	
3	1	Van Nha	
4	2	Thien Ha	
5	3	Ngan Giang	
6	4	Thuy Hang	
7	5	Thu Thuy	
8	6	Hanh Van	
9	7	Thao Nguyen	
10	8	Cong Son	
11	9	Pham Trung	
12	10	Thanh Tu	

	A	B	C	D
1	id	name	seat_capacity	
2	0	H1.100	30	
3	1	H1.101	30	
4	2	H1.102	35	
5	3	H1.103	35	
6	4	H1.200	35	
7	5	H1.201	40	
8	6	H2.100	40	
9	7	H2.101	45	
10	8	H2.102	45	
11	9	H2.103	45	
12	10	H2.200	50	
13	11	H2.201	50	
14	12	H3.100	55	

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.

3.2. 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.

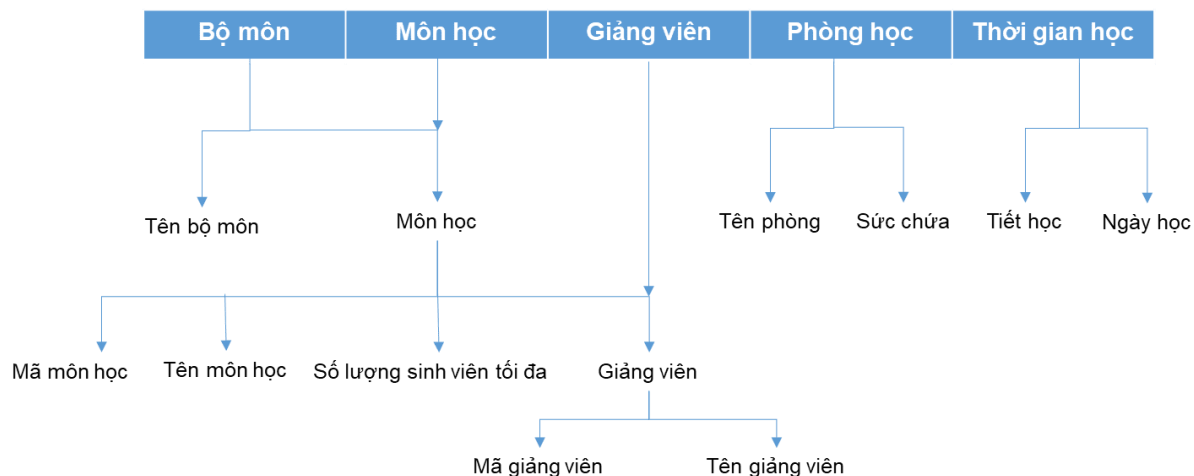
Thời gian học: Được chia thành hai thành phần:

- 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.

3.3. Thiết kế nhiệm 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.

3.4. 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:

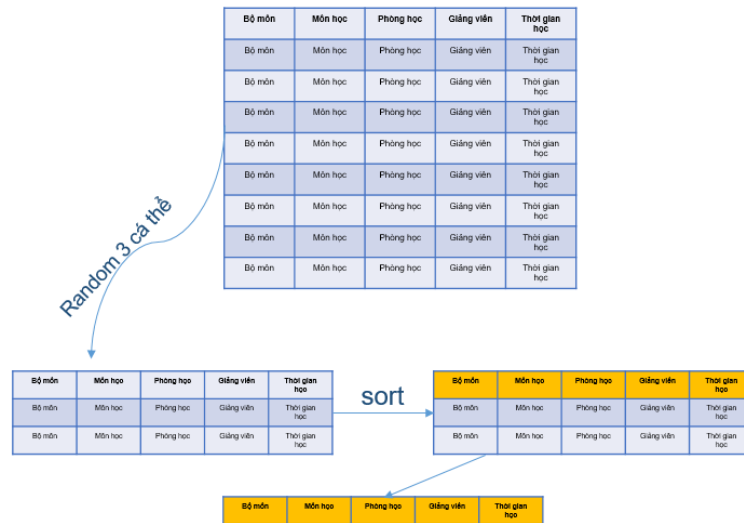
TOURNAMENT_SELECTION_SIZE = 3

Tỉ lệ đột biến sẽ xảy ra:

MUTATION_RATE = 0.1

3.4.1. Phép chọn cá thể 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.

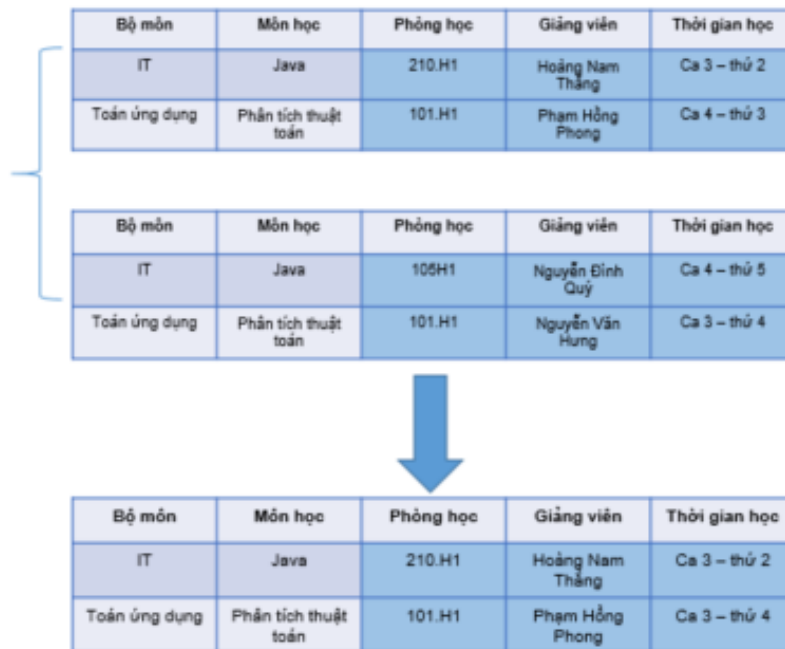


3.4.1. Phép lai ghép

Như đã nói ở trên phép lai ghép diễn ra bằng cách ghép một hàng 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.

3.4.2. 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

Bộ môn	Môn học	Phòng học	Giảng viên	Thời gian học
IT	Java	210.H1	Hoàng Nam Thắng	Ca 3 – thứ 2
Toán ứng dụng	Phân tích thuật toán	101.H1	Phạm Hồng Phong	Ca 3 – thứ 4



Bộ môn	Môn học	Phòng học	Giảng viên	Thời gian học
IT	Java	10.H3	Hoàng Nam Thắng	Ca 3 – thứ 2
Toán ứng dụng	Phân tích thuật toán	101.H1	Phạm Hồng Phong	Ca 3 – thứ 4

Sau đó thao tác đột biến tạo ra một lớp ngẫu nhiên và di chuyển nhiệm 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 nhiệm sắc thể

3.5. Độ thích nghi của hệ thống

Bộ môn	Môn học	Phòng học	Giảng viên	Thời gian học
IT	Java	210.H1	Hoàng Nam Thắng	Ca 3 – thứ 2
Toán ứng dụng	Phân tích thuật toán	101.H1	Phạm Hồng Phong	Ca 3 – thứ 4



Kiểm tra ràng buộc



$$\text{Độ thích nghi} = \frac{1}{1.0 * \text{conflict} + 1}$$

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 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.

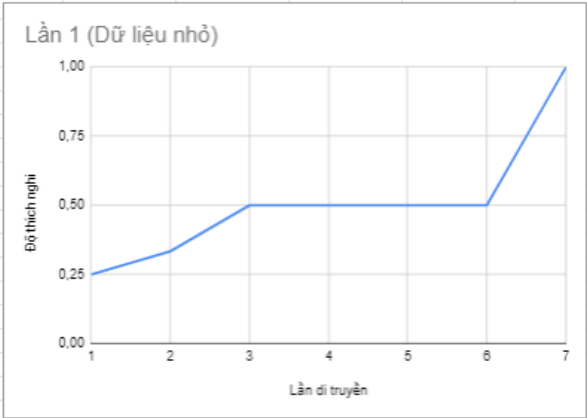
Chương 4. Các thực nghiệm của hệ thống

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](#)
- Hệ điều hành Windows 10 Pro.
- Compiler: Visual Studio code.

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	1	2	3	4	5	6	7
Độ thích nghi	0,25	0,33	0,50	0,50	0,50	0,50	1,00

Ca	Ngày	Phòng	Môn học	Giảng viên	Bộ môn
4-6	MON	H3.300	C++	Hanh Van	LANGUAGE
7-9	MON	H2.101	Caculus 2	Thien Ha	MATH
10-12	MON	H2.102	Badminton 1	So Khanh	PE
4-6	TUE	H2.200	Python	Thao Nguyen	IT
10-12	TUE	H2.100	Java	Thao Nguyen	IT
13-15	TUE	H2.103	Web	Thao Nguyen	IT
4-6	WED	H2.200	English 1	Ngan Giang	MATH
4-6	WED	H3.200	Table Tennis	Hao Nam	PE
7-9	WED	H1.103	Mordem Art 1	Thanh Tu	IT
13-15	WED	H2.200	Table Tennis	So Khanh	ART
13-15	WED	H3.200	Practical Anotomy	Hao Nam	PE
7-9	THU	H3.300	Mordem Art 2	Thanh Tu	ART

Chương 5. Kết luận và đề xuất hướng phát triển.

5.1. Kết quả đạt được

Áp dụng đọc 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.

5.2. 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 ưu 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.

Chương 6. 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