

**ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CNTT VÀ TRUYỀN THÔNG**

NGUYỄN THỊ HỒNG

**GIẢI THUẬT DI TRUYỀN VÀ ỨNG DỤNG VÀO
BÀI TOÁN LẬP THỜI KHÓA BIỂU**

LUẬN VĂN THẠC SỸ KHOA HỌC MÁY TÍNH

Thái Nguyên – 2013

**ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CNTT VÀ TRUYỀN THÔNG**

NGUYỄN THỊ HỒNG

**GIẢI THUẬT DI TRUYỀN VÀ ỨNG DỤNG VÀO
BÀI TOÁN LẬP THỜI KHÓA BIỂU**

Chuyên ngành: Khoa học máy tính
Mã số: 60 48 01

LUẬN VĂN THẠC SỸ KHOA HỌC MÁY TÍNH

**NGƯỜI HƯỚNG DẪN KHOA HỌC
TS. Lê Văn Phùng**

Thái Nguyên – 2013

MỤC LỤC

DANH MỤC CÁC KÝ HIỆU, CÁC TỪ VIẾT TẮT	3
DANH MỤC CÁC BẢNG.....	4
DANH MỤC CÁC HÌNH.....	5
LỜI MỞ ĐẦU	6
CHƯƠNG 1. GIẢI THUẬT DI TRUYỀN TRONG KHAI PHÁ DỮ LIỆU	8
1.1 Quá trình khai phá dữ liệu và giải thuật di truyền (GA).....	8
1.1.1 Quá trình khai phá dữ liệu.....	8
1.1.2 Giải thuật di truyền.....	11
1.2 Các khái niệm cơ bản về GA	12
1.2.1 Nhiễm sắc thể	12
1.2.2 Quần thể, thế hệ, toán tử di truyền, tiến hóa	14
1.2.3 Hàm thích nghi	14
1.2.4 Chọn lọc.....	14
1.2.5 Lai ghép.....	17
1.2.6 Đột biến	19
1.2.7 Chiến lược nạp lại quần thể.....	20
1.3 Mô hình GA	22
1.4 Không gian tìm kiếm và điều kiện dừng của GA	23
1.4.1 Không gian tìm kiếm.....	23
1.4.2 Điều kiện dừng của GA.....	24
1.5 Đặc điểm và ứng dụng của GA.....	24
1.5.1 Đặc điểm của GA	24
1.5.2 Ứng dụng của GA.....	25
CHƯƠNG 2. BÀI TOÁN THUỘC LỚP NP.....	26
VÀ BÀI TOÁN LẬP THỜI KHÓA BIỂU.....	26

2.1 Lớp các bài toán NP	26
2.1.1 Một số khái niệm:	26
2.1.2 NP-khó và NP-đầy đủ.....	26
2.1.3 Tối ưu hóa đa mục tiêu.....	27
2.2. Bài toán lập thời khóa biểu	28
2.2.1 Một số khái niệm cơ sở	32
2.2.2 Mô hình bài toán.....	33
CHƯƠNG 3. XÂY DỰNG BÀI TOÁN THỜI KHOÁ BIỂU DỰA TRÊN GIẢI	
THUẬT DI TRUYỀN	36
3.1 Bài toán lập thời khóa biểu ở trường Cao đẳng Công nghệ và Kinh tế Hà Nội .	36
3.1.1 Giới thiệu.....	36
3.1.2 Nội dung hoạt động nghiệp vụ	36
3.2 Áp dụng giải thuật di truyền vào bài toán lập thời khóa biểu.....	37
3.2.1 Các toán tử di truyền đối với bài toán hỗ trợ xếp thời khóa biểu.....	37
3.2.2 Thiết kế cơ sở dữ liệu	42
3.3. Cài đặt chương trình.....	49
3.4. Đánh giá kết quả thử nghiệm	49
KẾT LUẬN	52
TÀI LIỆU THAM KHẢO.....	53

DANH MỤC CÁC KÝ HIỆU, CÁC TỪ VIẾT TẮT

TT	Từ viết tắt	Tiếng Anh	Nghĩa Tiếng Việt
1	GA	Genetic Algorithm	giải thuật di truyền
2	EC	Evolutionary computation	tính toán tiến hóa
3	EP	Evolutionary Programming	quy hoạch tiến hóa
4	ES	Evolutionary Strategies	các chiến lược tiến hóa
5	GP	Genetic Programming	lập trình di truyền
6	CS	Classifier Systems	các hệ thống phân loại
7	MOOP	Multi-objective Optimization Problem	Tối ưu hóa đa mục tiêu
8	NST		nhiệm sắc thể
9		Selection	chọn lọc
10		Cross-over	lai ghép
11		Mutation	đột biến
12		Reproduction	sinh sản
13		pop-size	kích cỡ quần thể
14		NP-hard	bài toán NP khó
15		NP-complete	bài toán NP đầy đủ
16		Fitness	độ thích nghi

DANH MỤC CÁC BẢNG

Bảng1: Mã hóa nhị phân độ dài 20 bit.....	12
Bảng2: Mã hóa hoán vị 2 NST A&B.....	13
Bảng3: Mã hóa giá trị các NST A, B, C	13
Bảng4: Mặt nạ lai ghép đồng nhất	18
Bảng 5: Lai ghép một điểm cắt mã hóa hoán vị	19
Bảng 6: Phép đảo bit mã hóa nhị phân	19
Bảng 7: Hoán vị thứ tự mã hóa hoán vị	20
Bảng 8: Thay đổi giá trị trong mã hóa giá trị.....	20
Bảng 9: Bảng chỉ số	33
Bảng 10: Bảng kiểm tra ràng buộc về phòng.....	39
Bảng 11: Bảng kiểm tra ràng buộc về giảng viên.....	39
Bảng 12: Bảng kiểm tra ràng buộc về lớp sinh viên.....	40

DANH MỤC CÁC HÌNH

Hình 1: Xác suất của mỗi NST theo kiểu lựa chọn Roulette.....	15
Hình 2: Lựa chọn xếp hạng.....	16
Hình 3: Lai ghép một điểm cắt mã hóa nhị phân.....	17
Hình 4: Lai ghép hai điểm cắt mã hóa nhị phân.....	17
Hình 5: Lai ghép đồng nhất mã hóa nhị phân.....	18
Hình 6: Lai ghép số học mã hóa nhị phân.....	18
Hình 7: Chiến lược nạp lại hoàn toàn.....	20
Hình 8: Chiến lược nạp lại ngẫu nhiên.....	21
Hình 9: Chiến lược nạp lại theo mô hình cá thể ưu tú.....	21
Hình 10: Sơ đồ mô tả GA.....	22
Hình 11: Phân lớp tạm thời các bài toán.....	27
Hình 12: Ma trận cá thể.....	38
Hình 13: Lai ghép 1 điểm cắt.....	41
Hình 14: Sơ đồ giải thuật di truyền đề xuất.....	42
Hình 15: Nhập danh mục quản lý hình thức bài học.....	45
Hình 16: Quản lý danh mục loại phòng học.....	45
Hình 17: Quản lý danh mục lớp học.....	46
Hình 18: Quản lý danh mục môn học.....	46
Hình 19: Quản lý danh mục phòng học.....	47
Hình 20: Quản lý danh mục tòa nhà.....	47
Hình 21: Quản lý chương trình đào tạo.....	48
Hình 22: Cập nhật chương trình đào tạo khung.....	48
Hình 23: Cửa sổ xếp Thời khóa biểu.....	49

LỜI MỞ ĐẦU

Hiện nay trong ngành khoa học máy tính, việc tìm kiếm lời giải tối ưu cho các bài toán là vấn đề luôn được các nhà khoa học đặc biệt quan tâm. Mục đích chính của các thuật toán tìm kiếm lời giải là tìm ra lời giải tối ưu cho bài toán trong thời gian nhỏ nhất. Các thuật toán như tìm kiếm không có thông tin, vét cạn (tìm kiếm trên danh sách, trên cây hoặc đồ thị) hoặc các thuật toán tìm kiếm có thông tin được sử dụng nhiều trong không gian tìm kiếm nhỏ. Đối với không gian tìm kiếm lớn, việc tìm kiếm các lời giải tối ưu cho bài toán gặp nhiều khó khăn. Do đó, cần thiết phải có những thuật giải tốt và sử dụng kỹ thuật trí tuệ nhân tạo khi giải quyết các bài toán có không gian tìm kiếm lớn. Thuật giải di truyền (Genetic Algorithm - GA) là một trong những kỹ thuật tìm kiếm lời giải tối ưu đã đáp ứng được yêu cầu của nhiều bài toán và ứng dụng. Cùng với logic mờ, GA được ứng dụng rất rộng rãi trong các lĩnh vực phức tạp. Sự kết hợp giữa GA và logic mờ đã chứng tỏ được hiệu quả trong các vấn đề khó mà trước đây thường được giải quyết bằng các phương pháp thông thường hay các phương pháp cổ điển, nhất là trong các bài toán cần có sự lượng giá, đánh giá sự tối ưu của kết quả thu được. Chính vì vậy, GA đã trở thành một trong những đề tài nghiên cứu thu hút được nhiều sự quan tâm và hiện nay đã và đang đem đến rất nhiều ứng dụng trong thực tiễn.

Xuất phát từ thuyết tiến hóa muôn loài của Darwin, GA là một 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 trong những điều kiện được qui định sẵn của môi trường. GA là một thuật giải và mục tiêu của GA không nhằm đưa ra lời giải chính xác tối ưu mà là đưa ra lời giải tương đối tối ưu.

John Holland (1975) và Goldberg (1989) đã đề xuất và phát triển GA, là thuật giải tìm kiếm dựa trên cơ chế chọn lọc và di truyền tự nhiên. Thuật giải này sử dụng các nguyên lý di truyền về sự thích nghi và sự sống các cá thể thích nghi nhất trong tự nhiên.

Do tính hấp dẫn và tính thời sự của khai phá dữ liệu, đặc biệt là giải thuật di truyền, tôi đã chọn đề tài “***Giải thuật di truyền và ứng dụng vào bài toán lập thời***

khóa biểu” làm luận văn cao học của mình. Trong đó tập trung nghiên cứu các kỹ thuật lập lịch và chọn ra một kỹ thuật tiêu biểu để thực hiện bài toán thời khóa biểu phục vụ công tác giảng dạy của trường, nơi tôi đang công tác.

Chương 1. GIẢI THUẬT DI TRUYỀN TRONG KHAI PHÁ DỮ LIỆU

1.1 Quá trình khai phá dữ liệu và giải thuật di truyền (GA)

1.1.1 Quá trình khai phá dữ liệu

Theo Bách khoa toàn thư Việt Nam, tri thức là “*kết quả của quá trình nhận thức của con người về đối tượng được nhận thức, làm tái hiện trong tư tưởng của con người những thuộc tính, những mối quan hệ, những quy luật vận động, phát triển của đối tượng và diễn đạt bằng ngôn ngữ tự nhiên hay hệ thống ký hiệu khác*”.

Phát hiện tri thức là một quá trình bao gồm một dãy các bước lặp sau:

1. Làm sạch dữ liệu
2. Tích hợp dữ liệu
3. Chọn lựa dữ liệu
4. Chuyển đổi dữ liệu
5. Khai phá dữ liệu
6. Đánh giá các mẫu
7. Trình diễn tri thức

Trong đó, khai phá dữ liệu là bước quan trọng nhất trong tiến trình phát hiện tri thức.

Dữ liệu là những mô tả về sự vật, con người và sự kiện trong thế giới thực.

Dữ liệu bao gồm số, ký tự, văn bản, hình ảnh, đồ họa,... có một giá trị nào đó đối với người sử dụng và chúng được lưu trữ, xử lý trong máy tính.

Theo liệu bách khoa toàn thư, “khai phá dữ liệu” là khâu chủ yếu trong quá trình phát hiện tri thức từ dữ liệu để trợ giúp cho việc làm quyết định trong quản lý.