

ĐH Bách khoa Hà Nội

Ban Hà Bằng

**Các thuật toán đúng gần đúng giải bài toán
cực tiểu hóa độ trễ**

Chuyên ngành: Khoa học máy tính Mã số: 62480101

Nghiên cứu sinh: Ban Hà Bằng

Người hướng dẫn khoa học:

PGS.TS. Nguyễn Đức Nghĩa

2014

LỜI CAM ĐOAN

Tôi xin cam đoan luận án này là kết quả nghiên cứu của tôi. Các kết quả viết chung với các tác giả khác đều đã được sự nhất trí của các đồng tác giả khi đưa vào luận án. Những kiến thức tham khảo để hoàn thành luận án đều được trích dẫn đầy đủ từ danh mục tài liệu tham khảo.

Hà Nội, 04-2014

Người hướng dẫn khoa học

Tác giả luận án

PGS.TS. Nguyễn Đức Nghĩa

Ban Hà Bằng

LỜI CẢM ƠN

Luận án này được hoàn thành tại Bộ môn Khoa học Máy tính, Viện Công nghệ Thông tin và Truyền thông, Trường Đại học Bách Khoa Hà Nội dưới sự hướng dẫn của PGS. TS. Nguyễn Đức Nghĩa. Tôi xin chân thành cảm ơn Thầy hướng dẫn, người đã trực tiếp hướng dẫn khoa học và tận tình giúp đỡ tôi trong quá trình nghiên cứu.

Tôi xin bày tỏ lòng biết ơn tới Bố Mẹ và Gia đình đã giúp đỡ, tạo điều kiện cho tôi trong quá trình học tập và hoàn thành luận án này.

Tôi cũng xin gửi lời cảm ơn tới các thầy cô trong Viện Công Nghệ Thông Tin, cũng như các thầy cô trong trường Đại học Bách khoa Hà Nội đã truyền thụ những kiến thức bổ ích trong quá trình tôi học tập và nghiên cứu tại Trường.

Mặc dù đã rất cố gắng nhưng do thời thời gian và kiến thức còn hạn chế nên luận án chắc còn có nhiều thiếu sót. Tôi rất mong nhận được những ý kiến đóng góp quý báu từ các Thầy Cô và các bạn.

Mục Lục

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
Tóm tắt	v
Danh mục thuật ngữ	vii
Danh mục bảng	viii
Danh mục hình vẽ	x
CHƯƠNG 1 TỔNG QUAN VỀ BÀI TOÁN	1
1.1 Mô hình toán học của bài toán cực tiểu hóa độ trễ	2
1.2 Một số hướng tiếp cận giải bài toán tối ưu hóa tổ hợp	5
1.2.1 Thuật toán nhánh cận	5
1.2.2 Thuật toán di truyền	8
1.2.3 Thuật toán đàn kiến	10
1.2.4 Thuật toán Tabu	10
1.2.5 Thuật toán lân cận biến đổi	11
1.3 Các nghiên cứu liên quan giải bài toán MLP	11
1.3.1 Thuật toán đúng	12
1.3.2 Thuật toán gần đúng cận tỷ lệ	12
1.3.3 Thuật toán meta-heuristic	13
1.4 Mục đích, phạm vi nghiên cứu	14
1.5 Dữ liệu thực nghiệm	15
1.6 Kết quả của luận án	18
1.7 Cấu trúc của luận án	20
CHƯƠNG 2 THUẬT TOÁN NHÁNH CẬN	21
2.1 Lược đồ thuật toán	22
2.2 Kết quả thực nghiệm	26
2.2.1 Thực nghiệm bộ dữ liệu ngẫu nhiên	26
2.2.2 Thực nghiệm bộ dữ liệu thực	28
2.3 Kết luận chương 2	35
CHƯƠNG 3 CÁC THUẬT TOÁN GẦN ĐÚNG CẬN TỶ LỆ	36

3.1	Đánh giá thực nghiệm hiệu quả của các thuật toán gần đúng cận tỷ lệ	37
3.1.1	Các thuật toán gần đúng cận tỷ lệ	37
3.1.2	Kết quả thực nghiệm	45
3.2	Thuật toán dựa trên phương pháp Subgradient	53
3.2.1	Lược đồ thuật toán	53
3.2.2	Kết quả thực nghiệm	58
3.3	Kết luận chương 3.....	65
	CHƯƠNG 4 CÁC THUẬT TOÁN META-HEURISTIC.....	66
4.1.	Thuật toán di truyền	67
4.1.1	Lược đồ của thuật toán.....	67
4.1.2	Kết quả thực nghiệm	72
4.2	Thuật toán di truyền lai ghép đàn kiến.....	83
4.2.1	Lược đồ của thuật toán.....	84
4.2.2	Kết quả thực nghiệm	89
4.3	Thuật toán lai thuật toán Tabu và thuật toán lân cận biến đổi.....	96
4.3.1	Lược đồ của thuật toán.....	96
4.3.2	Kết quả thực nghiệm	104
4.4	Kết luận chương 4.....	112
	KẾT LUẬN	113
	DANH MỤC CÁC CÔNG TRÌNH	116
	CÔNG BỐ ĐƯỢC SỬ DỤNG TRONG LUẬN ÁN	116
	TÀI LIỆU THAM KHẢO	117

Tóm tắt

Bài toán cực tiểu hóa độ trễ (Minimum latency problem - MLP) dưới dạng tổng quát có thể phát biểu trong ngôn ngữ của lý thuyết đồ thị như sau: Cho $G = (V, E)$ là đồ thị vô hướng có trọng số không âm trên mỗi cạnh $e \in E$. Giả sử, T là một hành trình xuất phát từ đỉnh s , chúng ta định nghĩa độ trễ của một đỉnh v bất kỳ thuộc T là độ dài của đường đi từ đỉnh xuất phát s đến v trên T . Độ trễ của hành trình T được định nghĩa như là tổng độ trễ của tất cả các đỉnh thuộc hành trình T . Bài toán cực tiểu hóa độ trễ MLP yêu cầu tìm một hành trình T bắt đầu từ đỉnh xuất phát s đi qua tất cả các đỉnh còn lại của đồ thị với tổng độ trễ là nhỏ nhất.

Bài toán MLP có nhiều ứng dụng trong thực tiễn. Cụ thể, trong lý thuyết lập lịch khi một máy chủ hay một người thợ phải lên kế hoạch phục vụ một tập các yêu cầu sao cho tổng (trung bình) thời gian chờ đợi của các yêu cầu là cực tiểu. Trong tìm đường đi trên mạng, bài toán cũng được ứng dụng để tìm hành trình với tổng độ trễ là nhỏ nhất. Trong bài toán tìm kiếm thông tin, bài toán MLP được ứng dụng để cực tiểu hóa độ trễ của việc tìm kiếm thông tin trên mạng.

Mục đích nghiên cứu của chúng tôi trong luận án này là đề xuất các thuật toán giải bài toán MLP với chất lượng lời giải tốt hơn chất lượng lời giải của các thuật toán giải bài toán MLP đã được công bố. Đối với một bài toán NP-khó như bài toán MLP, hiện tại có ba hướng tiếp cận chính để phát triển thuật toán giải: 1) hướng tiếp cận đúng, 2) hướng tiếp cận gần đúng cận tỷ lệ, 3) hướng tiếp cận meta-heuristic. Đóng góp của chúng tôi trong luận án là đề xuất các thuật toán giải theo cả ba hướng tiếp cận:

- Phát triển thuật toán đúng đưa ra lời giải tối ưu cho bài toán MLP với kích thước bài toán lên đến 40 đỉnh.
- Khảo sát thực nghiệm về hiệu quả của các thuật toán gần đúng cận tỷ lệ hiện biết, là cơ sở để đề xuất thuật toán gần đúng mới có cận tỷ lệ tốt hơn.
- Phát triển ba thuật toán theo hướng tiếp cận meta-heuristic. Chúng tôi đề xuất thuật toán dựa trên lược đồ của thuật toán di truyền để giải bài toán MLP và một số kỹ thuật mới được tích hợp vào từng bước của thuật toán. Nhằm nâng cao chất lượng lời giải và thời gian chạy thuật toán, chúng tôi đề xuất hai thuật toán meta-heuristic lai là: Thuật toán (ACO-GA) lai ghép giữa thuật toán di truyền (GA) và thuật toán đàn kiến (ACO); và thuật toán TS-VNS lai ghép giữa thuật toán Tabu (TS) và thuật toán lân cận biến đổi (VNS).

Để đánh giá hiệu quả của các thuật toán đề xuất, chúng tôi tiến hành thực nghiệm trên các bộ dữ liệu chuẩn và so sánh kết quả thu được với kết quả của các công trình nghiên cứu liên quan. Kết quả thực nghiệm chỉ ra các thuật toán đề xuất đưa ra lời giải tốt hơn các thuật toán tốt nhất hiện biết trên nhiều bộ dữ liệu.

Danh mục thuật ngữ

STT	Từ viết tắt	Giải nghĩa tiếng Anh	Giải nghĩa tiếng Việt
1	ACO	Ant colony optimization	Tối ưu hoá đàn kiến
2	ACO-GA	-	Thuật toán di truyền lai ghép thuật toán đàn kiến
3	GA	Genetic algorithm	Thuật toán di truyền
4	TS	Tabu search	Tim kiếm Tabu
5	VNS	Variable neighborhood search	Tim kiếm lân cận biến đổi
6	TS-VNS	-	Thuật toán tabu lai ghép thuật toán đa lân cận
7	MLP	Minimum latency problem	Bài toán cực tiểu hóa độ trễ
8	TSP	Traveling salesman problem	Bài toán người du lịch
9	TRP	Traveling repairman problem	Bài toán thợ sửa chữa lưu động
10	DMP	Delivery man problem	Bài toán người giao hàng
11	TDTSP	Time dependent traveling Salesman problem	Bài toán người du lịch với thời gian bị chặn
12	DP	Dynamic programming	Quy hoạch động
13	B&B	Branch and bound	Phương pháp nhánh cận
14	CP	Constraint programming	Quy hoạch ràng buộc
15	-	Approximation algorithm	Thuật toán gần đúng
16	-	Simulated annealing algorithm	Thuật toán phỏng tôi luyện
17	-	Local search	Tim kiếm địa phương
18	GRASP	Greedy randomized adaptive search procedure	Thủ tục tìm kiếm tham lam ngẫu nhiên tự thích nghi
19	ILS	Iterated local search	Tim kiếm địa phương leo đồi
20	RVND	Random variable neighborhood descend	Tự lân cận biến đổi ngẫu nhiên
21	k -MST	k -minimum spanning tree	Bài toán cây khung nhỏ nhất đi qua k đỉnh
22	k -troll	Minimum k -troll problem	Bài toán hành trình ngắn nhất đi qua k đỉnh
23	PCST	Prize collecting steiner tree	Bài toán cây Steiner
24	-	Polynomial time algorithm (Polynomial algorithm)	Thuật toán thời gian tính đa thức
25		Benchmark test	Bộ dữ liệu chuẩn
26	<i>OPT</i>	Best known solution	Lời giải tốt nhất hiện biết
27	<i>SDT</i>	Social disaster technique	Kỹ thuật hủy diệt

Danh mục bảng

Bảng 1. 1 Mô tả các bộ dữ liệu	16
Bảng 2. 1 Thời gian chạy của thuật toán trong bộ dữ liệu ngẫu nhiên 1 (tính theo phút).....	30
Bảng 2. 2 Thời gian chạy của thuật toán trong bộ dữ liệu ngẫu nhiên 2 (tính theo phút).....	31
Bảng 2. 3 Thời gian chạy của thuật toán trong bộ dữ liệu thực 2 (tính theo phút)	32
Bảng 2. 4 Thời gian chạy của thuật toán trong bộ dữ liệu ngẫu nhiên 3 (TPR-10-Rx) (tính theo giây)	33
Bảng 2. 5 Thời gian chạy của thuật toán trong bộ dữ liệu ngẫu nhiên 3 (TPR-20-Rx) (tính theo giây)	34
Bảng 2. 6 Thời gian chạy của thuật toán cho các file dữ liệu nhỏ trong.....	34
Bảng 3. 1 Kết quả thực nghiệm các thuật toán trong các bộ dữ liệu nhỏ.....	49
Bảng 3. 2 Kết quả thực nghiệm các thuật toán trên bộ dữ liệu ngẫu nhiên 3 (TPR-50-Rx)	50
Bảng 3. 3 Kết quả thực nghiệm các thuật toán trên bộ dữ liệu ngẫu nhiên 3 (TPR-100-Rx)	51
Bảng 3. 4 Kết quả thực nghiệm các thuật toán trên bộ dữ liệu thực 1	52
Bảng 3. 5 Kết quả thực nghiệm cho các bộ dữ liệu nhỏ	61
Bảng 3. 6 Kết quả thực nghiệm với bộ dữ liệu ngẫu nhiên 3 (TPR-50-Rx)	62
Bảng 3. 7 Mô tả $\overline{gap}_1, \overline{gap}_2$ đối với các bộ dữ liệu nhỏ	62
Bảng 3. 8 Mô tả \bar{T} đối với các bộ dữ liệu nhỏ.....	62
Bảng 3. 9 Kết quả thực nghiệm với bộ dữ liệu ngẫu nhiên 3 (TPR-100-Rx)	63
Bảng 3. 10 Kết quả thực nghiệm với bộ dữ liệu thực 1	64
Bảng 3.11 Mô tả $\overline{gap}_1, \overline{gap}_2$ đối với các bộ dữ liệu lớn.....	64
Bảng 3. 12 Mô tả \bar{T} đối với các bộ dữ liệu lớn.....	64
Bảng 4. 1 Thực nghiệm lựa chọn kích thước quần thể	74
Bảng 4. 2 Thực nghiệm lựa chọn tham số xác suất lai ghép và đột biến.....	74
Bảng 4. 3 Thực nghiệm lựa chọn kích thước nhóm.....	74
Bảng 4. 4 Thực nghiệm lựa chọn tỷ lệ ζ_1	75
Bảng 4. 5 Thực nghiệm xác định giá trị NGD	75
Bảng 4. 6 Thực nghiệm xác định giá trị NGT	75
Bảng 4. 7 Kết quả thực nghiệm các thuật toán cho các bộ dữ liệu nhỏ	79
Bảng 4. 8 Kết quả thực nghiệm các thuật toán cho bộ dữ liệu ngẫu nhiên 3 (TPR-50-Rx).....	80
Bảng 4. 9 Kết quả thực nghiệm các thuật toán cho bộ dữ liệu ngẫu nhiên 3 (TPR-100-Rx).....	80
Bảng 4. 10 Kết quả thực nghiệm các thuật toán cho bộ dữ liệu ngẫu nhiên 3 (TPR-200-Rx).....	81
Bảng 4. 11 Kết quả thực nghiệm các thuật toán cho bộ dữ liệu thực 1.....	82
Bảng 4. 12 Mô tả \bar{T} theo phút đối với các bộ dữ liệu lớn.....	82
Bảng 4. 13 Kết quả thực nghiệm của các thuật toán.....	90
Bảng 4. 14 Kết quả so sánh các thuật toán cho cho các bộ dữ liệu nhỏ.....	92
Bảng 4. 15 Kết quả so sánh các thuật toán cho bộ dữ liệu ngẫu nhiên 3 (TPR-50-Rx).....	93
Bảng 4. 16 Kết quả so sánh các thuật toán cho bộ dữ liệu ngẫu nhiên 3 (TPR-100-Rx).....	93

Bảng 4. 17 Kết quả so sánh các thuật toán cho bộ dữ liệu ngẫu nhiên 3 (TPR-200-Rx).....	94
Bảng 4. 18 Kết quả so sánh các thuật toán cho bộ dữ liệu thực 2.....	95
Bảng 4. 19 Mô tả \bar{T} đối với các bộ dữ liệu lớn.....	95
Bảng 4. 20 Kết quả thực nghiệm các thuật toán cho các bộ dữ liệu nhỏ	107
Bảng 4. 21 Kết quả thực nghiệm các thuật toán cho các bộ dữ liệu ngẫu nhiên 3 (TPR-50-Rx)	108
Bảng 4. 22 Kết quả thực nghiệm các thuật toán cho bộ dữ liệu ngẫu nhiên 3 (TPR-100-Rx).....	108
Bảng 4. 23 Kết quả thực nghiệm các thuật toán cho các bộ dữ liệu ngẫu nhiên 3 (TPR-200-Rx)	109
Bảng 4. 24 Kết quả thực nghiệm các thuật toán cho các bộ dữ liệu ngẫu nhiên 3 (TPR-500-Rx)	109
Bảng 4. 25 Kết quả thực nghiệm các thuật toán cho bộ dữ liệu thực 1.....	110
Bảng 4. 26 So sánh độ lệch chuẩn độ trễ lời giải của các thuật toán	111
Bảng 4. 27 Mô tả \bar{T} đối với các bộ dữ liệu lớn.....	111
Bảng 5. 1 Tổng hợp các thuật toán đề xuất.....	115