

# CHỦ ĐỀ 1: GIẢI QUYẾT VẤN ĐỀ BẰNG TÌM KIẾM

## 1. Nội dung lý thuyết

### 1.1. Định nghĩa bài toán tìm kiếm

- Bài toán tìm kiếm là gì? Các thành phần của bài toán tìm kiếm?
- Qui đổi một số vấn đề thực tế thành bài toán tìm kiếm

### 1.2. Nhóm thuật toán tìm kiếm mù

#### 1.2.1. Tìm kiếm theo chiều rộng

- Ý tưởng của các thuật toán: **Breadth-First Search (BFS)**, **Least Cost Breadth-First Search (LCBFS)**, **Uniformed-Cost Search (UCS)**
- Điểm khác biệt cơ bản của các thuật toán này là gì? (tiêu chí chọn đỉnh kế tiếp, điều kiện dừng, ...)
- **Qui tắc tính chi phí đường đi  $g$**
- Cách sử dụng hàng đợi ưu tiên trong bài toán tìm kiếm
- Tính đầy đủ và tối ưu của các thuật toán (\*)

#### 1.2.2. Tìm kiếm theo chiều sâu

- Ý tưởng của các thuật toán: **Depth-First Search (DFS)**, **DFS cải tiến – PCDFS** và **MEMDFS**
- Tính đầy đủ và tối ưu của các thuật toán
- Điểm khác biệt giữa PCDFS và MEMDFS, ưu điểm của mỗi phương pháp trong trường hợp cụ thể

#### 1.2.3. Tìm kiếm lặp sâu dần

- Ý tưởng của thuật toán **Iterative Deepening Search (IDS)**.
- Tính đầy đủ và tối ưu của IDS

### 1.3. Nhóm thuật toán tìm kiếm có heuristic

#### 1.3.1. Tìm kiếm tham lam tốt nhất đầu tiên và A\*

- Ý tưởng của các thuật toán: **Greedy Best First Search (GBFS)**, **A\***
- Điểm khác biệt cơ bản của các thuật toán này so với nhóm thuật toán tìm kiếm mù là gì? (tiêu chí chọn đỉnh kế tiếp, điều kiện dừng, ...)
- **Qui tắc tính giá trị heuristic  $h$ , đại lượng  $f = g + h$**
- Tính đầy đủ và tối ưu của các thuật toán (\*)

### 1.3.2. Tìm kiếm lặp sâu dần A\*

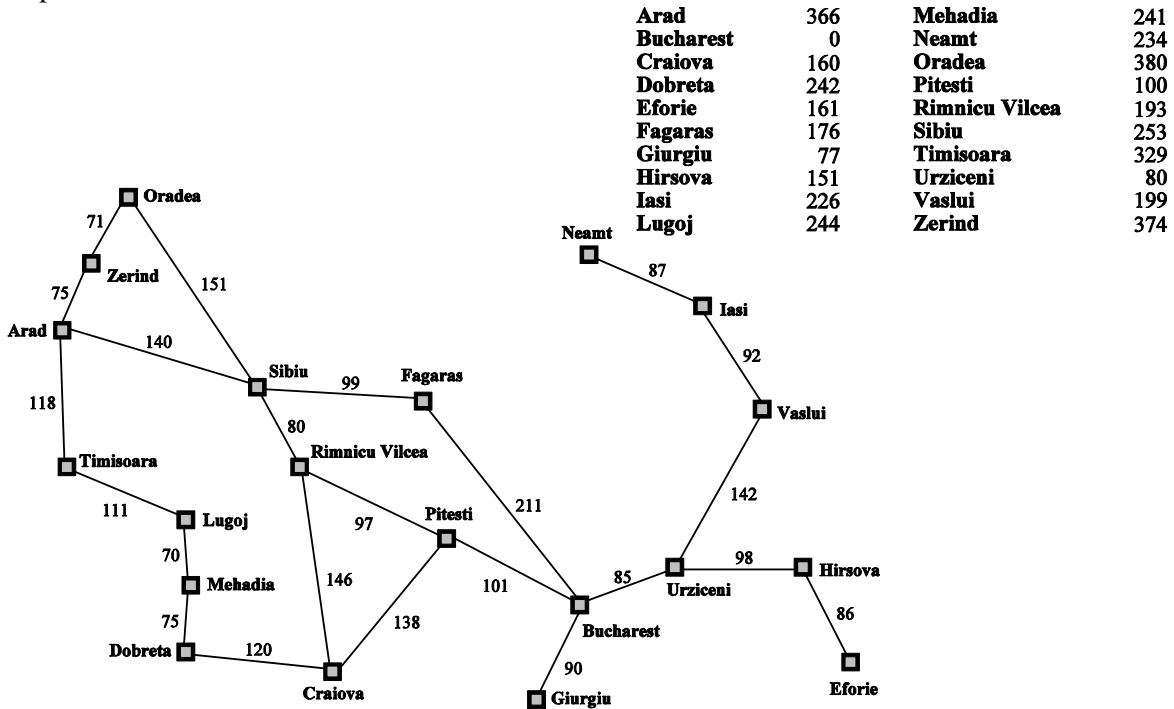
- Ý tưởng của thuật toán Iterative Deepening A\*
- Điểm khác biệt giữa IDS và IDA\*
- Tính đầy đủ và tối ưu của IDA\*

### 1.4. Thuật giải leo đồi và thuật giải di truyền

- Thuật giải leo đồi có đặc điểm gì giống và khác so với các thuật toán tìm kiếm mù và tìm kiếm heuristic? Có đảm bảo tìm thấy đường đi và đường đi tối ưu hay không? Trình bày một số cải tiến của thuật giải leo đồi.
- Ý tưởng của thuật giải di truyền: sơ đồ thuật giải, gen, các toán tử lai và đột biến, hàm thích nghi, hàm mục tiêu. Làm thế nào biểu diễn gen? Phương pháp bàn quay Roulette là gì?

## 2. Nội dung bài tập

2.1. Cho bản đồ các thành phố ở Rumani và khoảng cách đường chim bay từ các thành phố đến Bucharest như bên dưới.



Một khách du lịch muốn tìm đường đi từ Arad đến Bucharest.

- Hãy tìm đường đi theo từng chiến lược tìm kiếm dưới đây. Trình bày thứ tự mở các trạng thái, đường đi kết quả và chi phí.
  - LCBFS: để tiết kiệm thời gian, giả sử đường đi kết thúc tại Bucharest (không có đường đi đến Giurgiu và Urziceni) và Lugoj (không có đường đi đến Mehadia)
  - UCS
  - Greedy Best First Search: sử dụng heuristic là khoảng cách đường chim bay
  - A\*: sử dụng heuristic như trong GBFS
- Hãy tìm đường đi sao cho qua ít thành phố nhất. Trình bày thứ tự mở các trạng thái, đường đi kết quả và chi phí thực tế.
- Liệt kê 3 đường đi tùy chọn khi sử dụng thuật toán DFS (có kiểm tra trạng thái đang nằm trên đường đi). Trình bày thứ tự mở các trạng thái, đường đi kết quả và chi phí thực tế.
- Biểu diễn cây tìm kiếm cho từng chiến lược tìm kiếm như trên.

2.2. Cho bản đồ một số thành phố ở Châu Âu như bên dưới. Con số nằm trên đường nối giữa hai thành phố biểu thị thời gian lái xe trung bình (giờ) giữa cặp thành phố này.



Một người đi công tác muốn lái xe từ Warsaw đến Rome. Với mỗi chiến lược tìm kiếm dưới đây, hãy trình bày thứ tự mở các trạng thái, đường đi kết quả và thời gian lái. Trong mọi yêu cầu, nếu xảy ra tình trạng trạng thái có chi phí bằng nhau thì chọn mở trạng thái nào có tên nhỏ hơn theo thứ tự bảng chữ cái (ví dụ Budapest < Munich).

- Tìm kiếm theo chiều sâu (sử dụng chiến lược kiểm tra trạng thái đang nằm trên đường đi để tránh lặp vô tận)
- Tìm kiếm theo chiều rộng
- Tìm kiếm chi phí đồng nhất
- Tìm kiếm tham lam tốt nhất đầu tiên với heuristic:  $h(\text{Odesa}) = 20$  giờ,  $h(\text{Budapest}) = 12$  giờ,  $h(\text{Munich}) = 3$  giờ,  $h(\text{Venice}) = 3$  giờ,  $h(\text{Rome}) = 0$  giờ,  $h(\text{Warsaw}) = 30$  giờ.
- Tìm kiếm A\* với cùng heuristic như câu d.
- Heuristic trong câu d có chấp nhận được hay không? Hãy chứng minh.

2.3. Cho trạng thái đầu (a) và trạng thái đích (b) như bên dưới.

1		3
4	2	5
7	8	6

(a)

1	2	3
4	5	6
7	8	

(b)

Hãy sử dụng thuật toán A\* để biến đổi từ trạng thái (a) sang trạng thái (b) sao cho số ô cần đây là ít nhất. Heuristic được sử dụng lần lượt là *Khoảng cách Manhattan* và *Số ô sai so với trạng thái đích*. Với mỗi trường hợp, trình bày cây tìm kiếm và bộ giá trị (f, g, h).

2.4. Cho trạng thái đầu (a) và trạng thái đích (b) như bên dưới.

1	3	6
8	7	2
5		4

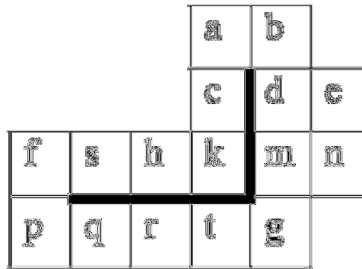
(a)

1	6	2
7	3	4
8	5	

(b)

Hãy sử dụng thuật toán A\* để biến đổi từ trạng thái (a) sang trạng thái (b) sao cho số ô cần đây là ít nhất. Heuristic được sử dụng lần lượt là *Khoảng cách Manhattan* và *Số ô sai so với trạng thái đích*. Với mỗi trường hợp, trình bày cây tìm kiếm và bộ giá trị (f, g, h).

2.5. Cho mê cung như hình bên dưới. Đường in đậm biểu diễn vách ngăn không qua được.



Hãy tìm đường đi từ s đến g với các chiến lược tìm kiếm dưới đây. Trình bày thứ tự duyệt các ô theo định dạng  $\langle b_1, b_2, \dots, b_n \rangle$ , với  $b_i$  là ô được duyệt.

- Tìm kiếm theo chiều rộng
- Tìm kiếm theo chiều sâu có kiểm tra trạng thái đang nằm trên đường đi để tránh lặp vô tận. Thứ tự mở là Phải  $\rightarrow$  Dưới  $\rightarrow$  Trái  $\rightarrow$  Trên.
- Tìm kiếm tham lam tốt nhất đầu tiên với heuristic là *khoảng cách Manhattan*.  $h(\text{state}) = \text{số bước ngắn nhất từ state đến g nếu không có rào chắn}$ , ví dụ,  $h(k) = 2$ ,  $h(s) = 4$ ,  $h(g) = 0$ .
- Tìm kiếm A\* với cách dừng thông thường

2.6. Cho mê cung như hình bên dưới. Đường in đậm biểu diễn vách ngăn không qua được.

M	N	O	<i>goal</i> P
I	J	K	L
E	F	G	H
<i>start</i> A	B	C	D

Hãy tìm đường đi từ **start** đến **goal** với các chiến lược tìm kiếm dưới đây. Trình bày thứ tự duyệt các ô theo định dạng  $\langle b_1, b_2, \dots, b_n \rangle$ , với  $b_i$  là ô được duyệt.

- a. Tìm kiếm theo chiều rộng.
- b. Tìm kiếm theo chiều sâu có kiểm tra trạng thái đang nằm trên đường đi để tránh lặp vô tận. Thứ tự mở là Phải  $\rightarrow$  Trái  $\rightarrow$  Trên  $\rightarrow$  Dưới.

## **CHỦ ĐỀ 2: BÀI TOÁN THỎA MÃN RÀNG BUỘC**

### **1. Nội dung lý thuyết**

#### **1.1. Bài toán người bán hàng/người du lịch**

- Dữ kiện được cung cấp và yêu cầu của bài toán
- Thuật toán GTS1 và GTS2
- Tính tối ưu của thuật toán: tối ưu cục bộ không đảm bảo tối ưu toàn cục.

#### **1.2. Bài toán phân công công việc**

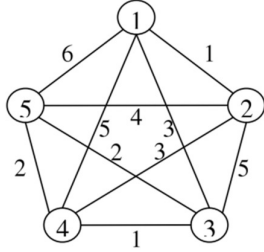
- Dữ kiện được cung cấp và yêu cầu của bài toán
- Nguyên lý sắp thứ tự để xây dựng lời giải
- Tính tối ưu của phương pháp giải quyết: không tối ưu

#### **1.3. Bài toán tô màu**

- Dữ kiện được cung cấp và yêu cầu của bài toán
- Các thuật toán tô màu heuristic: tô màu theo bậc và tô màu tham lam
- Điểm khác biệt cơ bản giữa hai thuật toán (cách chọn đỉnh để tô màu, lời giải, ...)

## 2. Nội dung bài tập

2.1. Cho ma trận kề biểu diễn chi phí đường đi giữa 5 thành phố. Tìm chu trình ngắn nhất qua 5 thành phố bằng thuật toán GTS2 với  $P = 3$ .



$\infty$	1	3	5	6
1	$\infty$	5	3	4
3	5	$\infty$	1	2
5	3	1	$\infty$	2
6	4	2	2	$\infty$

2.2. Cho ma trận kề biểu diễn chi phí đường đi giữa 7 thành phố. Tìm chu trình ngắn nhất qua 7 thành phố bằng thuật toán GTS2 với  $P = 4$ . (Lưu ý giá trị chi phí giữa các thành phố).

$\infty$	7	2	6	1	9	5
9	$\infty$	5	6	1	8	2
4	7	$\infty$	8	5	9	7
8	8	10	$\infty$	6	6	7
3	3	7	8	$\infty$	1	4
11	10	11	8	3	$\infty$	5
7	4	9	9	6	7	$\infty$

2.3. Phân công công việc cho 2 máy  $M_1, M_2$  và 5 công việc với thời gian thực hiện như sau  $T_1 = 3, T_2 = 3, T_3 = 2, T_4 = 2, T_5 = 2$ .

2.4. Một dịch vụ in ấn luận văn tốt nghiệp có 3 nhân viên đánh máy và một quản lý. Dịch vụ nhận được yêu cầu đánh máy luận văn của sinh viên tốt nghiệp như sau.

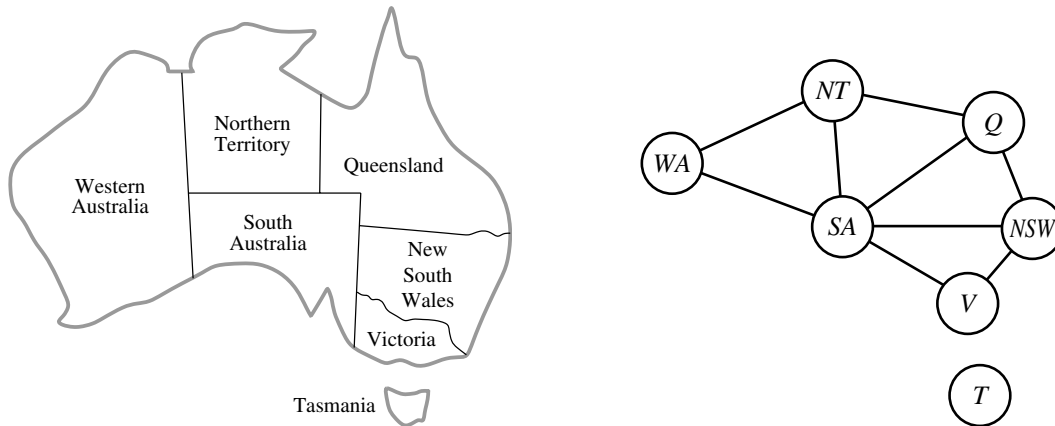
Luận văn	$L_1$	$L_2$	$L_3$	$L_4$	$L_5$	$L_6$	$L_7$	$L_8$	$L_9$	$L_{10}$	$L_{11}$	$L_{12}$
Số trang	200	140	70	100	60	120	50	80	100	150	40	60

Biết rằng trong 1 giờ, một nhân viên có thể đánh máy được 10 trang.

- Phân chia các luận văn cho 3 nhân viên đánh máy sao cho thời gian hoàn thành công việc là sớm nhất.
- Trong trường hợp người quản lý của tham gia đánh máy, công suất của người quản lý chỉ bằng  $\frac{1}{2}$  công suất của một nhân viên. Tìm cách chia các luận văn cho 3 nhân viên và người quản lý sao cho thời gian hoàn thành việc đánh máy luận văn là sớm nhất.



**2.5.** Cho bản đồ các bang của Australia như bên dưới. Hãy lần lượt áp dụng thuật toán tô màu theo bậc và tô màu tham lam để tô bản đồ sao cho các bang có chung biên giới không được tô trùng màu và số màu sử dụng là ít nhất.



**2.6.** Một bảng thi đấu bóng đá có 6 đội bóng. Biết rằng:

- Đội A đã đấu với đội B và C
- Đội B đã đấu với đội D và F
- Đội E đã đấu với đội C và F

Mỗi đội chỉ thi đấu 1 trận trong 1 tuần với đội khác. Hãy lập lịch thi đấu sao cho các trận còn lại sẽ được thực hiện trong số tuần ít nhất.

# CHỦ ĐỀ 4: CÁC THUẬT TOÁN HỌC MÁY

## 1. Nội dung lý thuyết

### 1.1. Giới thiệu Học máy

- Học máy là gì?
- Các phương pháp học: học có giám sát, học không có giám sát, học tăng cường

### 1.2. Phương pháp học cây quyết định

- Cấu trúc cây quyết định.
- Các độ đo chọn lựa thuộc tính tốt nhất: Entropy, Information Gain, Information Gain Ratio, Gini Index.
- Phương pháp rút luật IF – THEN từ cây quyết định
- Lưu ý khi xây dựng cây
  - Nếu nhánh có dữ liệu chưa phân hóa mà không còn thuộc tính kiểm tra thì tạo nút lá với giá trị  $x$  sao cho  $x$  là phân lớp chiếm đa số trong dữ liệu tại nhánh đó (nếu các phân lớp bằng nhau thì chọn ngẫu nhiên).
  - Trường hợp mẫu mới thiếu dữ liệu: nếu dữ liệu thiếu là thuộc tính kiểm tra thì kết luận là **Không xác định**.

### 1.3. Phương pháp thống kê xác suất Naïve Bayes

- Công thức xác suất Bayes. Giá trị xác suất  $P(Y = v)$ ,  $P(X_i = u | Y = v)$ .
- Công thức làm trơn Laplace. Lưu ý: áp dụng cho mọi giá trị xác suất
- Trường hợp mẫu mới thiếu dữ liệu: không xét thuộc tính bị thiếu giá trị vào tích xác suất.

### 1.4. Phương pháp học qui nạp

- Các bước của thuật toán ILA.
- Lưu ý khi tạo luật:
  - Chỉ tăng  $m$  khi không còn tổ hợp giá trị nào để tạo luật. Nếu  $m$  đã tăng bằng số lượng thuộc tính và không thể tạo tổ hợp giá trị thì bỏ qua và chuyển sang bảng khác.
  - Trường hợp mẫu mới thiếu dữ liệu: chỉ chọn luật nào thỏa toàn bộ tiền đề.

## 2. Nội dung bài tập

### 2.1. Bảng dữ liệu sau có thuộc tính phân lớp là Go Skiing?

Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết?
1	Âm	Sương mù	Vắng	Tốt	Không
2	Khô	Nắng	Vắng	Bị thương	Không
3	Khô	Nắng	Vắng	Tốt	Có
4	Khô	Nắng	Cao điểm	Tốt	Có
5	Khô	Nắng	Vừa phải	Tốt	Có
6	Băng	Gió	Cao điểm	Mệt mỏi	Không
7	Âm	Nắng	Vắng	Tốt	Có
8	Băng	Sương mù	Vừa phải	Tốt	Không
9	Khô	Gió	Vắng	Tốt	Có
10	Khô	Gió	Vắng	Tốt	Có
11	Khô	Sương mù	Vắng	Tốt	Có
12	Khô	Sương mù	Vắng	Tốt	Có
13	Âm	Nắng	Vừa phải	Tốt	Có
14	Băng	Sương mù	Vắng	Bị thương	Không
15	Âm	Gió	Cao điểm	Mệt mỏi	?
16	Khô	Sương mù	Vừa phải	Bị thương	?

### 2.2. Bảng dữ liệu sau có thuộc tính phân lớp là Kết quả?

Mẫu	Thời gian	Loại giải	Mặt sân	Dùng hết sức	Kết quả
1	Sáng	Master	Cỏ	1	F
2	Chiều	Grand slam	Đất nện	1	F
3	Tối	Giao hữu	Cứng	0	F
4	Chiều	Giao hữu	Thảm	0	N
5	Chiều	Master	Đất nện	1	N
6	Chiều	Grand slam	Cỏ	1	F
7	Chiều	Grand slam	Cứng	1	F
8	Chiều	Grand slam	Cứng	1	F
9	Sáng	Master	Cỏ	1	F
10	Chiều	Grand slam	Đất nện	1	N
11	Tối	Giao hữu	Cứng	0	F
12	Tối	Master	Thảm	1	N
13	Chiều	Master	Đất nện	1	N
14	Chiều	Master	Cỏ	1	F
15	Chiều	Grand slam	Cứng	1	F
16	Chiều	Grand slam	Đất nện	1	F
17	Sáng	Giao hữu	Thảm	0	?
18	Chiều	Master	Cỏ	0	?

2.3. Bảng dữ liệu sau có thuộc tính phân lớp là Đợi hay không?

Mẫu	Có đôi không?	Số khách	Loại món ăn	Đợi hay không?
1	Có	Một ít	Pháp	Đợi
2	Không	Một ít	Bánh mì	Đợi
3	Không	Đông khách	Pháp	Không
4	Có	Một ít	Ý	Đợi
5	Không	Không có	Bánh mì	Không
6	Không	Đông khách	Bánh mì	Không
7	Có	Đông khách	Ý	Không
8	Có	Đông khách	Bánh mì	Đợi
9	Có	Đông khách	Pháp	?
10	Không	Một ít	Ý	?

2.4. Bảng dữ liệu sau có thuộc tính phân lớp là Hoạt động.

Mẫu	Thời tiết	Ba mẹ ở nhà?	Tình trạng kinh tế	Hoạt động
1	Nắng	Có	Giàu	Xem phim
2	Nắng	Không	Giàu	Tennis
3	Gió	Có	Giàu	Xem phim
4	Mưa	Có	Nghèo	Xem phim
5	Mưa	Không	Giàu	Ở nhà
6	Mưa	Có	Nghèo	Xem phim
7	Gió	Không	Nghèo	Xem phim
8	Gió	Không	Giàu	Mua sắm
9	Gió	Có	Giàu	Xem phim
10	Nắng	Không	Giàu	Tennis
11	Gió	Có	Nghèo	?
12	Nắng	-	Giàu	?

2.5. Bảng dữ liệu sau có thuộc tính phân lớp là Mua máy tính?

Mẫu	Tuổi	Thu nhập	Sinh viên?	Tiết kiệm	Mua máy tính?
1	<= 30	cao	không	khá	Không
2	<= 30	cao	không	tốt	Không
3	31...40	cao	không	khá	Có
4	>40	trung bình	không	khá	Có
5	>40	thấp	không	khá	Có
6	>40	thấp	có	tốt	Không
7	31...40	thấp	có	tốt	Có
8	<=30	trung bình	không	khá	Không
9	<=30	thấp	có	khá	Có
10	>40	trung bình	có	khá	Có
11	<=30	trung bình	có	tốt	Có
12	31...40	trung bình	không	tốt	Có
13	31...40	cao	có	khá	Có
14	>40	trung bình	không	tốt	Không
15	31...40	trung bình	có	Khá	?
16	>40	cao	có	tốt	?