VIETNAM NATIONAL UNIVERSITY HO CHI MINH CITY
UNIVERSITY OF INFORMATION TECHNOLOGY

BÀI TẬP KIỂM TRA TÍNH ĐÚNG ĐẮN ĐO HIỆU NĂNG CỦA CHƯƠNG TRÌNH

CS112.P11.KHTN

Sinh Viên :
Hoàng Minh Thái - 23521414
Nguyễn Trọng Tất Thành

Giảng viên :
Nguyễn Thanh Sơn

Ngày 21 tháng 11 năm 2024

Mục lục

Chương 1

Bài 1: Bài toán Set Cover

1.1 Mô tả bài toán

Cho một tập U, là một tập hợp các phần tử, và một tập hợp $S = \{S_1, S_2, \ldots, S_m\}$ gồm các tập con của U. Mỗi tập con S_i là một tập con của U, và mục tiêu là chọn một số ít các tập con sao cho tất cả các phần tử trong U đều được bao phủ, tức là mỗi phần tử của U phải xuất hiện ít nhất một lần trong các tập con đã chọn.
Cụ thể:

- Tập $U = \{u_1, u_2, \ldots, u_n\}$ là tập hợp các phần tử cần được bao phủ.
- Tập con $S = \{S_1, S_2, \ldots, S_m\}$ là tập hợp các tập con của U, mỗi tập con S_i chứa một số phần tử của U.
- **Mục tiêu:** Chọn một số ít các tập con từ S sao cho mỗi phần tử trong U xuất hiện ít nhất một lần trong các tập con được chọn. Tối thiểu hóa số lượng tập con được chọn.

Định nghĩa chính thức:
Cho:

$$U = \{u_1, u_2, \ldots, u_n\}, \qquad S = \{S_1, S_2, \ldots, S_m\} \qquad \text{với } S_i \subseteq U \text{ cho mọi } i.$$

Mục tiêu là tìm một tập con các tập con $S' \subseteq S$ sao cho:

$$\bigcup_{S_i \in S'} S_i = U$$

và tối thiểu hóa số lượng các tập con được chọn, tức là:

$$\min |S'|.$$

1.2 Cách giải

1.2.1 Thuật toán tham lam

1. Chọn tập con $S_i \in S$ sao cho tập này bao phủ được nhiều phần tử chưa được bao

phủ nhất trong U .

2. Thêm S_i vào tập lời giải C.

3. Loại bỏ các phần tử đã được S_i bao phủ khỏi U .

Phân tích tỉ lệ xấp xỉ: Trong bài này, chi phí để chọn ra một tập được coi là như nhau
⇒ chọn tập có nhiều phần tử nhất sẽ lợi về chi phí nhất.

Giả sử OPT là chi phí của lời giải tối ưu. Trước một bước cụ thể của thuật toán tham lam, giả sử đã có $k-1$ phần tử được bao phủ. Tại bước này, chi phí để bao phủ phần tử thứ k thỏa mãn:

$$\text{Cost}(k) \leq \frac{\text{OPT}}{n-k+1}.$$

Điều này là do phần tử thứ k chưa được bao phủ, nghĩa là tồn tại một tập S_i (trong lời giải tối ưu) chứa phần tử này và chưa được chọn bởi thuật toán tham lam. Vì thuật toán tham lam chọn tập S_i hiệu quả nhất về chi phí, chi phí trên mỗi phần tử trong tập được chọn bởi thuật toán tham lam phải nhỏ hơn hoặc bằng chi phí trên mỗi phần tử trong lời giải tối ưu. Do đó:

$$\text{Cost}(k) = \frac{\text{OPT}}{|U-I|},$$

trong đó $|U-I|$ là số phần tử chưa được bao phủ bởi thuật toán tham lam.

Giá trị của $|U-I|$ bằng $n-(k-1)$, và điều này rút gọn thành $n-k+1$. Chi phí tổng thể của thuật toán tham lam là tổng của chi phí cho tất cả n phần tử, cụ thể:

$$\text{Cost}_{\text{Greedy}} \leq \frac{\text{OPT}}{n} + \frac{\text{OPT}}{n-1} + \ldots + \frac{\text{OPT}}{1}.$$

$$\leq \text{OPT} \cdot \left(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}\right).$$

Áp dụng bất đẳng thfic:

Greedy Cost

Tổng $1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}$ xấp xỉ $\ln(n)$. Do đó:

$$\text{Cost}_{\text{Greedy}} \leq \text{OPT} \cdot \ln(n).$$

Code

1.2.2 Giải pháp lập trình phi tuyến tính

Chúng ta định nghĩa một biến nhị phân x_i cho mỗi tập con S_i trong S, với:

- $x_i = 1$: nếu tập S_i được chọn vào lời giải.
- $x_i = 0$: nếu tập S_i không được chọn.

Hàm mục tiêu:

$$\min \sum_{i=1}^{m} x_i$$

Ràng buộc:

$$\sum_{i:u_j \in S_i} x_i \geq 1, \quad \forall u_j \in U,$$

$$x_i \in \{0, 1\}, \quad \forall i.$$

Trong đó:

- $\sum_{i:u_j \in S_i} x_i$: đảm bảo rằng mọi phần tử u_j đều được bao phủ bởi ít nhất một tập con S_i.
- $x_i \in \{0, 1\}$: biểu thị rằng chỉ có thể chọn hoặc không chọn một tập S_i.

Thư giãn bài toán LP: Để giải bài toán trong thời gian đa thức, ta thư giãn ràng buộc $x_i \in \{0, 1\}$ thành $x_i \in [0, 1]$.

Mô hình LP thư giãn:

- **Hàm mục tiêu:**

$$\min \sum_{i=1}^{m} x_i$$

- **Ràng buộc:**

$$\sum_{i:u_j \in S_i} x_i \geq 1, \quad \forall u_j \in U,$$

$$x_i \in [0, 1], \quad \forall i.$$

Bài toán LP này cho phép x_i nhận giá trị thực trong khoảng từ 0 đến 1.
Giải bài toán: Sử dụng các phương pháp tối ưu hóa tiêu chuẩn như:

- Thuật toán *Simplex*.
- *Interior Point*.

để tìm nghiệm tối ưu x^*_i của bài toán LP thư giãn.
Vì x^*_i có thể là giá trị phân số, ta cần chuyển đổi nghiệm này thành một nghiệm nguyên.

Code

1.2.3 Thuật toán gần đúng

Với mỗi tập S_i, nếu $x_i^* \geq \frac{1}{H(n)}$, chọn S_i vào lời giải. Trong đó, $H(n)$ là số Harmonic của n, được tính như sau:

$$H(n) = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} \approx \ln(n).$$

Phân tích tỉ lệ xấp xỉ:

Sau khi làm tròn, tất cả các phần tử trong U đều được bao phủ vì:

- Với mỗi $u_j \in U$, bài toán LP đảm bảo rằng $\sum_{i:u_j \in S_i} x_i^* \geq 1$.
- Bằng cách làm tròn với ngưỡng $\frac{1}{H(n)}$, mọi phần tử u_j vẫn được bao phủ.

Mỗi x_i^* được làm tròn lên tối đa $H(n)$ lần. Nên, tổng chi phí của lời giải làm tròn là:

$$\sum_{i=1}^{m} x_i \leq H(n) \cdot \sum_{i=1}^{m} x_i^*.$$

Nghiệm LP thư giãn cung cấp cận dưới cho lời giải tối ưu (OPT):

$$\sum_{i=1}^{m} x_i^* \leq \text{OPT}.$$

Vậy:

$$\text{Chi phí lời giải làm tròn} \leq H(n) \cdot \text{OPT}.$$

Vì $H(n) \approx \ln(n)$, thuật toán đạt tỉ lệ xấp xỉ $\ln(n)$.

Code

Chương 2

Bài 2: Bài toán TSP (Travelling Salesman Problem)

2.1 Cách giải

2.1.1 Thuật toán tham lam

Ý tưởng

Ưu điểm

Nhược điểm

Thuật toán

Code

2.1.2 Thuật toán Heuristic Nearest Neighbor (NNH)

Ý tưởng

Ưu điểm

Nhược điểm

Thuật toán

Code