Дискретная оптимизация
весна 2013

Александр Дайняк

http://www.dainiak.com

Задача о покрытии

- Задана $0,1$-матрица без нулевых столбцов.
- Строка матрицы *покрывает* столбец, если на их пересечении стоит «1».
- *Покрытие* матрицы — это подмножество строк, покрывающее все столбцы.
- Задача о покрытии — это задача отыскания покрытия, имеющего минимально возможную мощность.
- Задача о взвешенном покрытии — когда строкам матрицы приписаны веса и нужно найти покрытие минимального веса.

$$\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

Примеры задач о покрытии

- Выбрать минимальное число вопросов на экзамене, охватывающее все нужные темы.
- Показать рекламный ролик в социальной сети минимально возможному числу участников группы, но так, чтобы слухи о нём распространились во всей группе.
- Набрать команду специалистов для решения задачи, требующей знания нескольких областей, так, чтобы расходы на зарплату были минимальны.

Жадный алгоритм построения покрытия матрицы

Для матрицы M рассмотрим алгоритм:

1. $R \coloneqq \emptyset$
2. $C \coloneqq \text{столбцы, не покрытые строками из } R$
3. if $|C| > 0$:
4. $\quad r^* \coloneqq \underset{r \text{ — строка } M}{\operatorname{argmax}} \#\{\text{столбцы из } C, \text{покрываемые } r\}$
5. $\quad R \coloneqq R \cup \{r^*\}$
6. $\quad$ goto 2.
7. R — искомое покрытие матрицы

То есть на каждом шаге добавляем в R любую из строк, покрывающих наибольшее число из ещё не покрытых столбцов.

Соотношение мощностей жадного и оптимального покрытий

Рассмотрим задачу о покрытии без весов.

Мощность оптимального покрытия матрицы M обозначим $\tau(M)$.

Теорема. (Johnson'1974, Lovász'1975, Stein'1974)
Пусть M — произвольная матрица, в каждой строке которой не более k единиц.

Тогда покрытие M, которое строит жадный алгоритм, имеет мощность не более

$$(1 + \ln k) \cdot \tau(M)$$

Соотношение мощностей жадного и оптимального покрытий

Доказательство:

Пусть на матрице M был запущен ж.а., и он построил покрытие мощности t.

Припишем вес каждому *столбцу* матрицы M:

Пусть на некотором шаге алгоритма выбрана строка, покрывающая z не покрытых ранее столбцов.
Тогда каждому из этих столбцов припишем вес $\frac{1}{z}$.

Заметим, что

$$\sum_{s\text{ — столбец } M} \text{вес}(s) = t$$

Соотношение мощностей жадного и оптимального покрытий

Пусть T — оптимальное покрытие матрицы M.

Пусть r — произвольная строка из T.

Пусть $\{s_l, s_{l-1}, \dots, s_1\}$ — все столбцы, покрываемые r.

Каждый из этих столбцов, рано или поздно, покрывается жадным алгоритмом.
Будем считать, что они покрываются ж.а. *именно в таком порядке*:

$$s_l, s_{l-1} \dots, s_1$$

Перед итерацией ж.а., на которой покрывается столбец s_i,
в матрице M остаются непокрытыми по крайней мере i столбцов:

$$s_i, \dots, s_1$$

Соотношение мощностей жадного и оптимального покрытий

Пусть r — строка из оптимального покрытия.

Пусть $\{s_l, s_{l-1}, \dots, s_1\}$ — все столбцы, покрываемые r.

Перед итерацией ж.а., на которой покрывается столбец s_i, в матрице M остаются непокрытыми по крайней мере i столбцов.

Ж.а. выберет на этой итерации строку, покрывающую по крайней мере i не покрытых ранее столбцов (не хуже чем r).

Значит, вес, приписанный нами столбцу s_i, не превосходит $1/i$.

Соотношение мощностей жадного и оптимального покрытий

Пусть r — произвольная строка из оптимального покрытия T.

Пусть $\{s_1, s_2, \dots, s_l\}$ — все столбцы, покрываемые r.

Имеем

$$\sum_{i=1}^{l} \text{вес}(s_i) \leq \sum_{i=1}^{l} \frac{1}{i} \leq 1 + \ln l \leq 1 + \ln k\,,$$

где k — максимальное число единиц в строках матрицы M.

Соотношение мощностей жадного и оптимального покрытий

Для произвольной строки r из оптимального покрытия T имеем

$$\sum_{s\text{ покрывается } r} \text{вес}(s) \leq 1 + \ln k$$

Так как каждый столбец покрывается хотя бы одной строкой из T, то

$$t = \sum_{s\text{ — столбец } M} \text{вес}(s) \leq \sum_{r\in T} \sum_{s\text{ покр. } r} \text{вес}(s) \leq$$

$$\leq \tau(M) \cdot (1 + \ln k)$$

Соотношение мощностей жадного и оптимального покрытий

Теорема.
Пусть M — произвольная матрица, в каждой строке которой не более k единиц. Тогда покрытие, построенное ж.а., имеет размер не более $(1 + \ln k) \cdot \tau(M)$.

Теорема.
Для любого $k \geq 2$ существует матрица M,
в каждой строке которой не более k единиц, а покрытие, построенное для M с помощью ж.а., имеет размер не менее

$$\frac{(\log_2 k) - 1}{2} \cdot \tau(M)$$

Соотношение мощностей жадного и оптимального покрытий

Доказательство:

Для $a \in \boldsymbol{N}$ рассмотрим матрицу M размера $(a+2) \times (2^{a+1}-2)$, у которой

- для $i = 1, \ldots, a$ каждая из строк r_i содержит ровно 2^i единиц,
- ни у какой пары строк из $\{r_1, \ldots, r_a\}$ нет общих единичных позиций,
- строка r_{a+1} содержит $(2^a - 1)$ единиц, и для каждого $i = 1, \ldots, a$ у r_{a+1} и r_i ровно 2^{i-1} общих единичных позиций,
- строка r_{a+2} — побитовое отрицание r_{a+1}.

Соотношение мощностей жадного и оптимального покрытий

На матрице M ж.а. будет работать так:

- Сначала ж.а. выберет строку r_a, покрыв 2^a столбцов.
- Строки r_{a+1} и r_{a+2} могут покрыть по
$$(2^a - 1) - 2^{a-1} = 2^{a-1} - 1$$
новых столбцов, поэтому r_{a-1} у них выиграет, и ж.а. выберет r_{a-1}.
- На следующем шаге ж.а. выберет r_{a-2} и т.д.

В итоге ж.а. выберет строки
$$r_a, r_{a-1}, \dots, r_1$$

Соотношение мощностей жадного и оптимального покрытий

В матрице M ж.а. выберет строки $r_a, r_{a-1}, \dots, r_1$.

При этом $\tau(M) = 2$, так как оптимальное покрытие: $\{r_{a+1}, r_{a+2}\}$.

Положим $a \coloneqq \lfloor \log_2 k \rfloor$.
Для такого a

- в каждой строке M не более $2^a \leq k$ единиц,
- жадное покрытие хуже оптимального в $\frac{a}{2} \geq \frac{(\log_2 k) - 1}{2}$ раз.

Жадный алгоритм во взвешенной задаче о покрытии

1. $R \coloneqq \emptyset$
2. $C \coloneqq \text{столбцы, не покрытые строками из } R$
3. if $|C| > 0$:
4. $\quad r^* \coloneqq \underset{r\text{ — строка } M}{\operatorname{argmin}} \frac{\text{вес строки } r}{\#\{\text{столбцы из } C\text{,покрываемые } r\}}$
5. $\quad R \coloneqq R \cup \{r^*\}$
6. $\quad$ goto 2.
7. R — искомое покрытие матрицы

То есть на каждом шаге добавляем в R любую из строк, минимизирующих «стоимость покрытия в расчёте на один столбец».

Соотношение мощностей жадного и оптимального покрытий

Вес оптимального покрытия матрицы M обозначим $\tau(M)$.

Теорема-упражнение. (Chvátal ’1979)
Пусть M — произвольная матрица, в каждой строке которой не более k единиц.

Тогда вес покрытия M, которое строит жадный алгоритм, не превосходит

$$(1 + \ln k) \cdot \tau(M)$$

Соотношение мощностей жадного и оптимального покрытий

Вес оптимального покрытия матрицы M обозначим $\tau(M)$.

Теорема.
Пусть M — произвольная матрица с n столбцами, каждой строке которой приписан вес из интервала $(0,1]$.

Тогда вес покрытия M, которое строит жадный алгоритм, не превосходит

$$\left(1 + \ln \frac{n}{\tau(M)}\right) \cdot \tau(M) + 1$$

Соотношение мощностей жадного и оптимального покрытий

Пусть $s_n, s_{n-1} \dots, s_1$ — столбцы, в порядке, в котором они покрываются жадным алгоритмом.

Перед итерацией ж.а., на которой покрывается столбец x_l, в матрице остаются непокрытыми l столбцов. Их все можно покрыть множеством строк с суммарным весом не более чем $\tau(M)$.

Значит, ж.а. на этой итерации выберет строку, вес которой в расчёте на один покрываемый столбец не больше $\frac{l}{\tau(M)}$.

Значит, для каждого l общая стоимость покрытия столбцов $x_n, \dots, x_l$ не превосходит $\sum_{i=l}^{n} \frac{\tau(M)}{i} = \tau(M) \cdot \sum_{i=l}^{n} \frac{1}{i} \leq \tau(M) \ln(n/l)$.

А стоимость покрытия столбцов $s_{l-1}, \dots, s_1$ не превосходит $(l-1)$.

Соотношение мощностей жадного и оптимального покрытий

Значит, для каждого l общая стоимость покрытия столбцов $x_n, \ldots, x_l$ не превосходит $\tau(M)\ln(n/l)$.

Стоимость покрытия столбцов $s_{l-1}, \ldots, s_1$ не превосходит $(l-1)$.

Значит, общая стоимость жадного покрытия не больше

$$\tau(M)\ln(n/l) + (l-1)$$

Взяв $l \coloneqq 1 + \lceil \tau(M) \rceil$, получаем, что стоимость жадного покрытия

$$\leq \left(1 + \ln\frac{n}{\tau(M)}\right) \cdot \tau(M) + 1$$