

组合优化理论

第5章 背包问题

主讲教师, 陈安龙



第5章 背包问题

- § 1 背包问题的描述
- § 2 分支定界法解背包问题
- § 3 近似算法解背包问题
- § 4 背包问题的一些相关问题



背包问题(Knapsack Problem)是一个有着广泛应用的组合优化问题,它不仅在投资决策、装载、库存等方面有应用,而且常以子问题形式出现在大规模优化问题中,它的理论与算法具有一定的代表性.

§ 1 背包问题的描述

背包问题的一般描述为:设有物品集 $U = \{u_1, u_2, ..., u_n\}$ 是一个准备放入容量为 $C \in \mathbb{Z}^+$ 的背包中的 n 项物品的集合.

如何选择 U 中的一些物品装入背包,使这些物品的总重量不超过 C,且使总价值达到最大?

背包问题的数学模型:

$$\max \quad z = \sum_{j=1}^{n} p_{j} x_{j}$$

$$(KP) \qquad \sum_{j=1}^{n} w_j x_j \le C$$

$$w_j \in \mathbb{Z}$$
— 重量

$$p_i \in \mathbb{Z}$$
——价值

$$C \in Z^+$$
——容量

$$x_{j} = 0$$
 or 1 $j \in N = \{1, 2, ..., n\}$

因为决策变量 $x_i = 0$ or 1,所以也称 0-1 背包问题.

一般背包问题:
$$x_j \in \mathbb{Z}^+ \cup \{0\}$$
 $j \in \mathbb{N} = \{1, 2, ..., n\}$

(General Knapsack Problem)



为讨论方便,总可假定(相当于标准化):

1,
$$w_j > 0$$
, $p_j > 0$ $j = 1 \sim n$;

$$2, w_j \leq C \quad j = 1 \sim n;$$

$$3, \quad \sum_{j=1}^{n} w_j > C$$

4、
$$\frac{p_1}{w_1} \ge \frac{p_2}{w_2} \ge \cdots \ge \frac{p_n}{w_n}$$
 即按价值密度从大到小排列.

§ 1 背包问题的描述

对 4 只需 $O(n \ln n)$ 次运算即可;

对 3 若
$$\sum_{j=1}^{n} w_{j} \leq C$$
 ,最优解为 $x_{1} = x_{2} = \cdots = x_{n} = 1$;

对 2 若 $w_j > C$,则最优解中 $x_j = 0$, ... 事先可去掉;

对 1 分三种情况讨论:

(1) 若
$$p_j \le 0$$
 且 $w_j \ge 0$ 令 $N^0 = \{j \in N | p_j \le 0, w_j \ge 0\}$

此时,最优解中 $x_i = 0$ 所以,该物品事先可去掉;

此时,最优解中 $x_i = 1$ 所以,该物品事先可去掉;



(3) 若 $p_j < 0$ 且 $w_j < 0$ 令 $N^- = \{ j \in N \mid p_j < 0, w_j < 0 \}$

只需在模型中,令 $x_i = 1 - y_i$,则系数即为大于零了.

综上(1)(2)(3)的假定,可作如下处理,使之满足:

对
$$\forall j \in \mathbb{N}^- \Leftrightarrow x_j = 1 - y_j \quad \overline{p}_j = -p_j \quad \overline{w}_j = -w_j$$

对于 $\forall j \in N^+ = N \setminus (N^0 \cup N^1 \cup N^-)$

则原问题化为:

$$\max \ z = \sum_{j \in N^{-} \cup N^{+}} \overline{p}_{j} y_{j} + \sum_{j \in N^{-} \cup N^{1}} p_{j}$$

$$s.t. \ \sum_{j \in N^{-} \cup N^{+}} \overline{w}_{j} y_{j} \le C - \sum_{j \in N^{-} \cup N^{1}} w_{j}$$

$$y_{j} = 0 \quad or \quad 1, \qquad j \in N^{-} \cup N^{+}$$

§1 背包问题的描述

如果在(KP)中,令 $x_j = 1 - y_j$

max
$$z = \sum_{j=1}^{n} p_{j} - \sum_{j=1}^{n} p_{j} y_{j}$$

s.t.
$$\sum_{j=1}^{n} w_{j} - \sum_{j=1}^{n} w_{j} y_{j} \leq C$$

$$y_{j} = 0$$
 or 1 $j \in N = \{1, 2, ..., n\}$

至少有q的物品不能放进去

该问题的实际意义是求不

$$\max \quad z = \sum_{j=1}^{n} p_{j} x_{j}$$

$$(KP) \quad s.t. \quad \sum_{j=1}^{n} w_{j} x_{j} \leq C$$

$$x_{j} = 0 \quad or \quad 1$$

$$j \in N = \{1, 2, ..., n\}$$



§ 2 背包问题的分支定界法

分支定界法(Branch and Bound Method)的基本思想在算法课有介绍,它的重要在于它提出了一类新的思路,使得许多原来不好解决的问题有了解决的可能性。(具有普适性)

确定问题(子问题)的最优值的上(下)界

通常是通过求解松弛问题,用松弛问题的解作为界

Note: 松弛问题选择的原则

- 1、松弛问题要与原问题的最优值尽量接近;
- 2、松弛问题要尽量容易解.

这两个原则不易统一,所以可选择不同的松弛问题



考虑 KP 的松弛问题:

$$x_j \in \{0,1\} \Rightarrow x_j \in [0,1]$$

$$\max \quad z = \sum_{j=1}^{n} p_{j} x_{j}$$

$$C(KP) \qquad s.t. \qquad \sum_{j=1}^{n} w_j x_j \leq C$$

$$0 \le x_j \le 1 \ j \in N = \{1, 2, ..., n\}$$

思路:将物品按价值密度从大到小的顺序放入包内,.....

记
$$s = \min \left\{ j \mid \sum_{i=1}^{j} w_i > C \right\}$$
 — 关键项 第一个放不下的 物品的序号



§ 2 背包问题的分支定界法

Theorem 6.1 C(KP) 最优解为

$$\frac{-}{x_{j}} = 1 \quad j = 1 \sim s - 1 \qquad \frac{-}{x_{j}} = 0 \quad j = s + 1 \sim n$$

$$\frac{-}{x_{s}} = \frac{\overline{C}}{w_{s}} \qquad \qquad \text{其中} \quad \overline{C} = C - \sum_{j=1}^{s-1} w_{j}$$
最优解值为:
$$z_{opt}(C(KP)) = \sum_{j=1}^{s-1} p_{j} + \overline{C} \frac{p_{s}}{w_{s}}$$

显然, $z_{opt}(KP) \le z_{opt}(C(KP))$, 由于 p_j 的整数性,

得到z(KP)的一个上界:

$$U_{1} = \left\lfloor z_{opt}(C(KP)) \right\rfloor = \sum_{j=1}^{s-1} p_{j} + \left\lfloor \frac{C}{C} \frac{p_{s}}{w_{s}} \right\rfloor$$

 α 表示不超过 α 的最大整数.

证明: 显然 C(KP) 的最优解必满足 $\sum_{i=1}^{n} w_i x_j = C$

$$\sum_{j=1}^{n} w_j x_j = C$$

设 x^* 是其最优解,要证 $x^* = x$ 若存在k < s 使 $x_k^* < 1$

则至少存在 $q \ge s$ 使 $x_q^* > x_q$. 取充分小的 $\varepsilon > 0$

(满足: 1、
$$x_k^* + \varepsilon \le 1$$
 2、 $x_q^* - \varepsilon \frac{w_k}{w} \ge 0$)

 $x_k^* + \varepsilon$ 则重量增加 εw_k , $x_q^* - \varepsilon \frac{w_k}{w}$ 则重量减少 εw_k

将 x_k^* 增加 ε , x_q^* 减少 $\varepsilon \frac{w_k}{w}$, 此时,仍是一个可行解,

且目标函数值增加 $\varepsilon(p_k - p_q \frac{w_k}{w}) > 0$, 矛盾.

$$\therefore \forall \forall k < s, \quad x_k^* = x_k = 1$$
 同理可证 $\forall \forall k > s \quad x_k^* = 0$

又由极大性知: $x_s^* = \frac{C}{w} = \frac{x}{x_s}$ 因此, $\frac{1}{x}$ 是最优解.

易证 $z_{ont}(C(KP))$ 为最优解值

§ 2 背包问题的分支定界法

Theorem 6.2 读
$$U^0 = \sum_{j=1}^{s-1} p_j + \left[\frac{\overline{C}}{w_{s+1}} \right]$$

$$U^{1} = \sum_{j=1}^{s-1} p_{j} + \left[p_{s} - (w_{s} - \overline{C}) \frac{p_{s-1}}{w_{s-1}} \right] \qquad U_{1} = \left[z_{opt}(C(KP)) \right] = \sum_{j=1}^{s-1} p_{j} + \left[\overline{C} \frac{p_{s}}{w_{s}} \right]$$

其中 s 与 \overline{c} 定义同前. 则

- 1、z(KP) 的一个上界为 $U_2 = \max(U^0, U^1)$;
- 2、对背包问题任一实例, $U_2 \leq U_1$

作为上界 U_2 比 U_1 更好

$$\sum_{j=1}^{s-1} p_j + p_s - (w_s - \overline{C}) \frac{p_{s-1}}{w_{s-1}}$$

$$\frac{\int_{C} C}{C} = C - \sum_{i=1}^{s-1} w_{i}$$

 W_1

 $w_s - C$



Proof: 1、因为 KP 中 x_s 不能取分数, 所以, KP 的最优解一定是 $x_s = 0$ or $x_s = 1$ 情形之一.

当 $\overline{x}_s = 0$, 由 Th 6.1 可知, U^0 是此情形的上界;

当 $\bar{x}_s = 1$, 这时,若 $\bar{x}_1 = \bar{x}_2 = \cdots = \bar{x}_{s-1} = 1$

重量超出 $w_s - \overline{C}$,而此时价值密度值最小的是 $\frac{p_{s-1}}{w_{s-1}}$.

∴ U¹是此情形的上界.

从而 $U_2 = \max(U^0, U^1)$ 是 z(KP) 的上界.



§ 2 背包问题的分支定界法

 w_1 $w_s - \overline{C}$ w_s w_s $w_s - \overline{C}$ w_s w_s w

只需证: $U^0 \leq U_1$ and $U^1 \leq U_1$

$$U_1 = \sum_{j=1}^{s-1} p_j + \left[\frac{\overline{C}}{\overline{W}_s} \right]$$

$$U^0 = \sum_{j=1}^{s-1} p_j + \left[\frac{\overline{C}}{w_{s+1}} \right]$$

$$\therefore \overline{C} \frac{p_{s+1}}{w_{s+1}} \leq \overline{C} \frac{p_s}{w_s}$$

§ 2 背包问题的分支定界法

 $2 \cdot U^0 \leq U_1$ 是显然的;

$$C(KP)$$
 的最优解值
$$z_{opt}(C(KP)) = \sum_{j=1}^{s-1} p_j + \overline{C} \frac{p_s}{w_s}$$

$$= \sum_{j=1}^{s} p_j - (w_s - \overline{C}) \frac{p_s}{w_s}$$
 (*)

C(KP) 当 $\frac{1}{x_s} = 1$ 的最优解值

$$\sum_{j=1}^{s} p_{j} - (w_{s} - \overline{C}) \frac{p_{s-1}}{w_{s-1}}$$
 (**)

$$: w_{s} > \overline{C}, \quad \frac{p_{s-1}}{w_{s-1}} \ge \frac{p_{s}}{w_{s}} \quad : \quad \sum_{j=1}^{s} p_{j} - (w_{s} - \overline{C}) \frac{p_{s}}{w_{s}} \ge \sum_{j=1}^{s} p_{j} - (w_{s} - \overline{C}) \frac{p_{s-1}}{w_{s-1}}$$

$$\therefore \quad \boldsymbol{U}^1 \leq \boldsymbol{U}_1 \quad \text{从而 } \boldsymbol{U}_2 \leq \boldsymbol{U}_1.$$

一般给出的上界越小,计算量越大,但越容易被剪枝.

用分支定界法求如下 KP:

$$\therefore x_{opt} = (1, 0, 0, 1, 0, 0, 0) \qquad z_{max} = 107$$



通过前面介绍的 C(KP), 自然想到如下贪婪算法

(Greedy Algorithm):
$$x_1 = \cdots = x_{s-1} = 1$$
, $x_s = \cdots = x_n = 0$

其目标函数值为 $\sum_{j=1}^{s-1} p_j$.

$$GA_0$$
 step 1 $x_0 = w_0 = 0$ $k = 1$

step 2 若
$$\sum_{j=0}^{k-1} w_j x_j + w_k \le C$$
, 则 $x_k = 1$,

否则
$$x_k = 0$$
;

step 3 若
$$k=n$$
 ,则结束;

否则 k := k+1, 转 step 2.



构造例子 I: n=2, $p_1=1+\varepsilon$, $w_1=1$, $p_2=k$, $w_2=k$, C=k. ε 为充分小的正数 .

按上述算法, 得: $x_1 = 1$, $x_2 = 0$, $z_{GA_0}(I) = 1 + \varepsilon$;

而显然最优解为: $x_1 = 0$, $x_2 = 1$, $z_{opt}(I) = k$.

$$\frac{z_{GA_0}(I)}{z_{opt}(I)} = \frac{1+\varepsilon}{k} \to 0 \qquad (k \to +\infty)$$

说明 *GA*₀ 的绝对性能比不会大于任意给定的正数, 所以,它不能作为近似解.但稍加改进,就可得到一

个绝对性能比为常数的较好的近似算法.

已知 GA_0 对解 0-1 背包问题效果很差,但在一般背 包问题中却是可以的.

设 max
$$z = \sum_{j=1}^{n} p_j x_j$$

$$s.t. \sum_{j=1}^{n} w_j x_j \le C$$

$$x_j \in z^+ \cup \{0\}, j = 1 \sim n$$

显然,
$$x_1 = \left\lfloor \frac{C}{w_1} \right\rfloor$$
, $x_2 = \cdots = x_n = 0$

是一个可行解

$$\therefore z_{GA_0} \geq p_1 \left\lfloor \frac{C}{w_1} \right\rfloor$$

通过求解 C(KP) 得 $z_{opt}(I) \le p_1 \frac{C}{w_1}$ 最优解值

松弛问题的



Theorem 6.3

对于一般背包问题有 $R_{GA_0} = \frac{1}{2}$ 成立

证明:

$$\left| \frac{C}{w_1} \right| \ge 1 \Rightarrow 1 + \left| \frac{C}{w_1} \right| \ge 2 \Rightarrow \frac{1}{1 + \left| \frac{C}{w_1} \right|} \le \frac{1}{2}$$

$$\therefore \frac{z_{GA_0}(I)}{z_{opt}(I)} \ge \frac{\mathbf{p}_1 \left\lfloor \frac{C}{w_1} \right\rfloor}{\mathbf{p}_1 \times \frac{C}{w_1}} \ge \frac{\left\lfloor \frac{C}{w_1} \right\rfloor}{1 + \left\lfloor \frac{C}{w_1} \right\rfloor} = 1 - \frac{1}{1 + \left\lfloor \frac{C}{w_1} \right\rfloor} \ge \frac{1}{2}$$



构造例子 $I: \mathcal{E}$ 为充分小的正数.

$$p_1 = k + 3\varepsilon, w_1 = \frac{k}{2} + \varepsilon, \quad p_2 = k, w_2 = \frac{k}{2}; \quad C = k$$

$$z_{GA_0}(I) = k+3\varepsilon, x_1 = 1, x_2 = 0$$

$$z_{opt}(I) = 2k, \quad x_1 = 0, \quad x_2 = 1$$

$$\frac{z_{GA_0}(I)}{z_{ont}(I)} = \frac{k+3\varepsilon}{2k} \to \frac{1}{2} \quad (k \to \infty)$$

所以
$$R_{GA_0} = \frac{1}{2}$$



 GA_1 step 1 求解 C(KP), 得关键项记为 s;

step 2 取
$$\max \left\{ \sum_{j=1}^{s-1} p_j, p_s \right\}$$
 作为近似解值.

即若
$$\sum_{j=1}^{s-1} p_j > p_s$$
,则 $x_1 = \cdots = x_{s-1} = 1, x_s = \cdots = x_n = 0$

否则
$$x_1 = \cdots = x_{s-1} = x_{s+1} = \cdots = x_n = 0, x_s = 1$$

Example 2 用 GA_1 求如下 KP:

$$n = 4$$
 $(p_j) = (3,7,17,20), (w_j) = (1,3,8,10)$ $C = 11$

Solution: 易验证
$$\frac{3}{1} \ge \frac{7}{3} \ge \frac{17}{8} \ge \frac{20}{10}$$
 有改进的吗?

显然,物品3 为关键项(即 s=3) : $p_1 + p_2 = 3 + 7 = 10$,

而
$$p_3 = 17$$
 近似解为 $x_1 = x_2 = x_4 = 0, x_3 = 1; z_{GA_1} = 17$

© Combination Optimization PPT was designed by Chen Anlong, @UESTC.edu.cn

Theorem 6.4
$$R_{GA_1} = \frac{1}{2} \cdot R_{GA_1} = \sup_{I} \left\{ r \le 1 \middle| \frac{z_{GA_1}(I)}{z_{opt}(I)} \ge r \right\}$$

Proof:由s的定义知:对于任意实例 I

$$z_{opt}(I) \leq \sum_{j=1}^{s-1} p_j + p_s$$

$$\sum_{i=1}^{s-1} z_{GA_1}(I) = \max \left\{ \sum_{j=1}^{s-1} p_j, p_s \right\} \ge \frac{1}{2} \left(\sum_{j=1}^{s-1} p_j + p_s \right)$$

因此,
$$\frac{z_{GA_1}(I)}{z_{out}(I)} \ge \frac{1}{2}$$
 从而 $R_{GA_1} \ge \frac{1}{2}$

继续证明 Theorem 6.4

取实例 I: n = 3, $p_1 = 1 + \varepsilon$, $w_1 = 1$, ε 为充分小的正数.

$$p_2 = w_2 = p_3 = w_3 = k$$
, $C = 2k$
 $z_{GA_1}(I) = k+1+\varepsilon$, $w_1 + w_2 = k+1$, $x_1 = x_2 = 1$, $x_3 = 0$
 $z_{opt}(I) = 2k$, $x_1 = 0$, $x_2 = x_3 = 1$
 $\frac{z_{GA_1}(I)}{z_{opt}(I)} = \frac{k+1+\varepsilon}{2k} \to \frac{1}{2}$ $(k \to \infty)$
所以 $R_{GA_1} = \frac{1}{2}$



 GA_2 step 1 求解 C(KP), 得关键项记为 s;

$$step 2$$
 令 $p_i = \max_j \{p_j\}$ 若 $\sum_{j=1}^{s-1} p_j > p_i$ 则 $x_1 = \dots = x_{s-1} = 1, x_s = \dots = x_n = 0$ 否则 $x_1 = \dots = x_{i-1} = x_{i+1} = \dots = x_n = 0$, $x_i = 1$

Example 3 用 GA_2 求如下 KP:

$$n = 4$$
 $(p_j) = (3,7,17,20), (w_j) = (1,3,8,10)$ $C = 11$

Solution: : $p_1 + p_2 = 3 + 7 = 10$,

而
$$\max_{j} \{p_{j}\} = p_{4} = 20$$
 近似解为 $x_{1} = x_{2} = x_{3} = 0, x_{4} = 1;$ $z_{GA_{2}} = 20$



Theorem 6.5
$$R_{GA_2} = \frac{1}{2}$$
 证明与 Th 6.3 的类似. 对 Ex. 3

$$n = 4$$
 $(p_j) = (3,7,17,20), (w_j) = (1,3,8,10)$ $C = 11$

考虑对 GA_2 进行修改,进一步可取

$$x_1 = x_4 = 1, \quad x_2 = x_3 = 0 \quad \forall z = 23$$

这在实际计算中是会有好处的.但绝对性能比不会改进.



1975年 Sahni 给出一个多项式时间近似算法.

算法 S_k :

step 1 对任意满足 $M \subseteq N = \{1, 2, \dots, n\}, |M| \le k$

且 $\sum_{i \in M} w_i \leq C$ 的子集 M , 先将 M 中的物品放入包内,

然后用算法 GA_1 或 GA_2 求解一个如下定义的 KP:

物品集为 NM ,包容量为 $C - \sum_{i \in M} w_i$,将得到的解与

M 的并作为原问题的近似解.

step 2 取上述所有不同解中最好一个作为输出.



§ 4 0-1背包问题的一些相关问题

一、有界背包问题 (Bounded Knapsack Problem)

$$x_i = 0$$
 or 1;

$$x_i \in z^+ \cup \{0\}$$
 ; (无界背包问题)

$$0 \le x_j \le b_j$$
 b_j 为给定的正整数.

显然,
$$GKP$$
 可化为 BKP , 只需令 $b_j = \frac{C}{w_j}$

BKP 可化为等价的 0-1 KP

思想: 任一整数可用 0,1 变量来表示,如x<16非负整数

$$x = x_1 + 2x_2 + 2^2x_3 + 2^3x_4$$
 \$\text{ \$\text{II} 13 } x_1 = 1, x_2 = 0, x_3 = 1, x_4 = 1\$



有界背包问题的求解

与前面讨论一样,总可假定 p_j, w_j, b_j, C 都是正整数;

$$\sum_{j=1}^{n} b_{j} w_{j} > C, \quad b_{j} w_{j} \leq C, \quad \frac{p_{1}}{w_{1}} \geq \frac{p_{2}}{w_{2}} \geq \cdots \geq \frac{p_{n}}{w_{n}}.$$

$$Theorem 6.7 \quad \text{id} \quad s = \min \left\{ j \middle| \sum_{i=1}^{j} b_{i} w_{i} > C \right\}, \quad \overline{C} = C - \sum_{i=1}^{s-1} b_{j} w_{j},$$

则 (1)
$$U_1 = \sum_{j=1}^{s-1} b_j p_j + \left[\frac{\overline{C} p_s}{w_s} \right]$$
 是 *BKP* 的一个上界;

(2) 取 $x_j = b_j (j = 1, 2, \dots, s - 1)$ $x_j = 0 (j = s, \dots, n)$ 得到一个可行解,将此解的值与 $b_s p_s$ 的值比较,取大者为输出。

该算法的绝对性能比 $R = \frac{1}{2}$, 计算复杂性为 $O(n \log n)$.



二、子集和问题

在组合优化问题中,很多问题是相通的,参数稍作修正,可能就是另一个问题.

在 0-1 背包问题中,令 $p_j = w_j$ (这时 $\frac{p_j}{w_j} = 1$, $j = 1 \sim n$

即

$$\max \quad z = \sum_{j=1}^n w_j x_j$$

称为子集和问题

$$(SSP) \quad s.t. \quad \sum_{j=1}^{n} w_j x_j \le C$$

Subset Sum Problem

$$x_{j} = 0$$
 or 1 $j \in N = \{1, 2, ..., n\}$

SSP 是 0-1KP 的特殊情形,所以原方法皆可用. 但因其特殊,它应有更简单的方法.

冷自上协<u>从</u>工担当应

子集和问题的求解

1、贪婪算法(GA):

按任意顺序将物品逐个放入包内,直到放不下为止,然后将其结果与最大重量的物品比较,取优者为输出.

可证
$$R_{GA} = \frac{1}{2}$$
 计算复杂性为 $O(n)$.

2、近似算法 MTGS:

将上述的贪婪算法分别用于物品集 $\{1,2,\dots,n\}$,

$$\{2,3,\dots,n\},\{3,\dots,n\},\dots,\{n-1,n\},\{n\}$$
,取最好者为输出.

这里假定
$$w_1 \geq w_2 \geq \cdots \geq w_n$$
.

可证
$$R_{MTGS} = \frac{3}{4}$$
 计算复杂性为 $O(nlogn+n^2)=O(n^2)$.

2023年10月18日 Combination Optimization PPT was designed by Chen Anlong, @UESTC.edu.cn



本章结束