第 28 卷 第 1 期 2008 年 1 月
河北大学学报(自然科学版)
Journal of Hebei University(Natural Science Edition)
Vol. 28 No. 1 Jan. 2008

预算型最大覆盖问题的近似算法

张　生, 何尚录
(兰州交通大学 数理与软件工程学院, 甘肃 兰州 730070)

摘　要: 研究了给定预算常数的最大覆盖问题, 给出了求解此问题的改进贪婪算法, 得到了性能保证为 $1-e^{-1}$ 的近似算法.

关键词: 覆盖问题; 贪婪算法; 下模集函数; 性能保证
中图分类号: O 221.7　　文献标识码: A　　文章编号: 1000-1565(2008)01-0007-03

Approximate Algorithm for the Budgeted Maximum Coverage Problem

ZHANG Sheng, HE Shang-lu
(School of Mathematics, Physics & Software Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: Discuss the maximum coverage problem subject to a budgeted constant, presents an improved greedy algorithm which has $(1-e^{-1})$-performance guarantee for this problem.

Key words: coverage problem; greedy algorithm; submodular set function; performance guarantee

覆盖问题主要分为集合覆盖问题和最大覆盖问题两大类, 是一类在生产实践中有着广泛应用的组合优化问题. 文[1~3]中讨论了最大覆盖问题在选址问题中的具体应用. 另外, 最大覆盖问题在电路布置、设施定位等各方面也有广泛的应用[4].

所谓预算型最大覆盖问题是指: 给定集合 $\mathbf{X}=\{x_1, x_2, \cdots, x_n\}$, 对于任一 $x_i \in \mathbf{X}$ 都对应非负权 $w_i(i=1, 2, \cdots, n)$, 集族 $S=\{S_1, S_2, \cdots, S_m\}$, 其中 $\mathbf{S}_j \subseteq \mathbf{X}(j=1, 2, \cdots, m)$ 且对应有费用系数 $c_j(j=1, 2, \cdots, m)$, 给定总的预算费用 $L \in R^+$, 问题的目标是从 S 中找出子集族 $H \subseteq S$ 使得在 H 中所包含元素的总费用不超过 L 的前提下, 使 H 所覆盖 $\mathbf{X}$ 的元素的总权数达到最大. 令

$f(H)=\sum_{x_i \in H} w_i$ 表示 H 所覆盖 $\mathbf{X}$ 中元素的总权数. (H' 表示 H 所覆盖 $\mathbf{X}$ 的元素)

$F(H)=\sum_{S_j \in H} c_i$ 表示 H 所包含的元素的总费用.

预算型最大覆盖问题的数学模型如下:

$$\max_{H \subseteq S}\{f(H) \mid F(H) \leqslant L\}. \tag{1}$$

下面, 引入非减下模集函数的概念[5], 然后说明 $f(H)$ 即为非减下模集函数.

令 $\mathbf{N}$ 为一有限集, 2^N 表示由 $\mathbf{N}$ 的所有子集组成的集合, 若集函数 $f: 2^N \to \mathbf{R}$ 满足: 1) 对任意的集合 $U, V \subseteq \mathbf{N}$, 如果 $U \subseteq V$, 则必有 $f(U) \leqslant f(V)$; 2) 对任意的集合 $U, V \subseteq \mathbf{N}$, 总有 $f(U)+f(V) \geqslant f(U \cup V)+f(U \cap V)$, 则称 f 为 2^N 到 $\mathbf{R}$ 的非减下模集函数.

收稿日期: 2007-06-17
基金项目: 国家自然科学基金资助项目(604730304); 兰州交通大学"青蓝"工程资助项目(QL-03-19A)
作者简介: 张　生(1981-), 男, 内蒙古丰镇人, 在读硕士研究生.

下模集函数有下面的判定性质[6].

定义在 2^N 到 $\mathbf{R}$ 上的非减集函数 $f(H)$ 为下模集函数当且仅当对任意 $U, V \subseteq \mathbf{N}$, 总有

$$f(U) \leqslant f(V) + \sum_{S_i \in U \setminus V} [f(V \cup \{S_i\}) - f(V)]. \tag{2}$$

由 $f(H)$ 表示 H 所覆盖 $\mathbf{X}$ 中元素的总权数与权的非负性不难得出 $f(H)$ 为集族 S 到权集 $\{w_i\}_{i=1}^n$ 上的非减下模集函数.

在[7]中给出了求解单一约束条件下非减下模集函数最大值的贪婪近似算法, 其性能保证为 $1-e^{-1}$. 同样, 文[6]从另一角度考虑这一问题, 也得到了相同性能保证的近似算法. 文[8]中考虑的分组约束下最大覆盖问题, 文[9]中考虑的是文[8]中的一种特殊情况. 本文结合文[6,7]的思想给出预算型最大覆盖问题的改进贪婪近似算法, 其性能保证为 $1-e^{-1}$. 所谓性能保证是指若由近似算法所求的近似解总是精确解的 α 倍, 则称此近似算法的性能保证为 α.

1 算法设计及其性能分析

1.1 算法设计

首先, 给出求解预算型最大覆盖问题的改进贪婪算法. 先给出下面将要用到的符号的意义: $|U|$ 表示 U 中元素的个数; $f(U)$ 表示 U 所覆盖 $\mathbf{X}$ 中元素的总权数; $F(U)$ 表示 U 所包含的元素的总费用; U' 表示 U 所包含的元素(集合)的并集; $S \setminus U$ 表示 S 与 U 的差集. 下面具体描述这一算法的步骤及基本思想.

第 1 步: 找出任意满足 $U \subseteq S = \{S_1, S_2, \cdots, S_m\}$, $|U| < 3$ 且 $F(U) \leqslant L$ 的子集族, 并分别计算 $f(U)$, 令 $f(H_1) = \max\{f(U)\}$.

第 2 步: 算法考虑所有可行的 $U \subseteq S = \{S_1, S_2, \cdots, S_m\}$, $|U| = 3$ 并计算 $f(U)$, 然后将 S 中剩余的元素贪婪地加到 U 中并保证可行性. 令 $f(H_2) = \max\{f(U)\}$.

第 3 步: 比较 $f(H_1)$ 与 $f(H_2)$, 如果 $f(H_1) > f(H_2)$, 输出 H_1; 反之, 输出 H_2.

下面, 详细给出这一算法的程序设计步骤.

算法:

$H_1 \leftarrow \arg\max\{f(U) \mid U \subseteq S, |U| < 3, 且 F(U) \leqslant L\}$,

$H_2 \leftarrow \varnothing$

对所有 $U \subseteq S$, $|U| = 3$ 且 $F(U) \leqslant L$ 令

$t = 1$, $U^0 = U$, $S^0 = S \setminus U$.

选出集合 $S_i \in S^0$ 使得 $\dfrac{f(S^{t-1} \cup \{S_i\}) - f(S^{t-1})}{c_i}$ 取得最大值

若 $F(U^0) + c_i \leqslant L$, 则取 $U^t = U^{t-1} \cup \{S_i\}$; $S^t = S^{t-1}$.

反之, 取 $U^t = U^{t-1}$; $S^t = S^{t-1} \setminus \{S_i\}$,

$t = t + 1$,

直到 $S^t \setminus U^t = \varnothing$ 为止.

若 $f(U^t) > f(H_2)$ 则取 $H_2 \leftarrow U^t$. 若 $f(H_1) > f(H_2)$, 输出 H_1, 反之, 输出 H_2.

由于元素不超过 3 个的子集族至多 $\sum_{i=1}^{3} \begin{pmatrix} m \\ 3 \end{pmatrix} \leqslant 3m^3$ 个, 故算法至多循环 $3m^3$ 次, 每次循环需调用 1 次贪婪算法求解, 其复杂性可达到 $O(m)$, 故上述算法的时间复杂性为 $O(3m^4)$.

1.2 性能分析

定理 1 利用上述算法求解预算型最大覆盖问题, 求得近似解的性能保证为 $1-e^{-1}$.

证明 用 OPT 表示问题(1)的最优解, 不失一般性, 假设 $|OPT| > 3$, 否则, 上述算法可求得其最优解. 对最优解集族 OPT 中的元素重新排序使得

$$f(OPT)=\max_{S_i\in OPT\setminus\{S_1,\cdots,S_{i-1}\}}[f(\{S_1,\cdots,S_{i-1}\}\cup\{S_i\})-f(\{S_1,\cdots,S_{i-1}\})],$$

上式的意义即为在 OPT 中将元素重新排序后使得 S_1 为所有元素中总权数最大的元素, S_2 为与 S_1 结合后使得总权数增量最大的元素, 依此类推. 令 $V=\{S_1,S_2,S_3\}$ 表示重新排序后 OPT 中前 3 个元素组成的集合, 则对于任意元素 $S_k\in OPT$, $k\geqslant 4$ 与子集族 $Z\subseteq S\setminus\{S_1,S_2,S_3,S_k\}$, 利用下模集函数的定义与集族 OPT 中元素的有序性可得

$$f(V\cup Z\cup\{S_k\})-f(V\cup Z)\leqslant f(\{S_k\})\leqslant f(\{S_1\});$$

$$f(V\cup Z\cup\{S_k\})-f(V\cup Z)\leqslant f(\{S_1\}\cup\{S_k\})-f(\{S_1\})\leqslant f(\{S_1,S_2\})-f(\{S_1\});$$

$$f(V\cup Z\cup\{S_k\})-f(V\cup Z)\leqslant f(\{S_1,S_2\}\cup\{S_k\})-f(\{S_1,S_2\})\leqslant$$
$$f(\{S_1,S_2,S_3\})-f(\{S_1,S_2\});$$

将上面 3 个不等式相加得

$$3f(V\cup Z\cup\{S_k\})-3f(V\cup Z)\leqslant f(\{S_1,S_2,S_3\})-f(\{S_1,S_2\})+$$
$$f(\{S_1,S_2\})-f(\{S_1\})+f(\{S_1\})=f(V). \tag{3}$$

下面假设上述算法在第 2 步中已选取子集族 $V=\{S_1,S_2,S_3\}$, 令 $U^0=V=\{S_1,S_2,S_3\}$; 令 t 是算法直到第 1 个集合 $S_{t^*+1}\in OPT$ 却未被选中加入 U^{t^*} 时总共迭代的次数, 因为 S_{t^*+1} 的加入会超出预算常数 L 的限制, 即 $c_{t^*+1}+F(U^{t^*})=\overline{L}>L$. 下面定义集函数 $g(H)=f(H)-f(V)$, 因 $f(H)$ 为非减下模集函数, 不难得出 $g(H)$ 也为非减下模集函数. 令 $U^t(t=1,2,\cdots,t^*)$ 表示算法在第 t 次迭代后得到的近似解. 利用不等式(2) 与非减下模集函数 $g(H)$ 的定义可得

$$g(OPT)\leqslant g(U^t)+\sum_{S_i\in OPT\setminus U^t}(g(U^t\cup\{S_i\})-g(U^t))=$$
$$g(U^t)+\sum_{S_i\in OPT\setminus U^t}(f(U^t\cup\{S_i\})-f(U^t))\leqslant g(U^t)+L\theta_{t+1}, \tag{4}$$

对于所有的 $t=1,2,\cdots,t^*$ 都成立. 这里 $\theta_{t+1}=\max\limits_{S_i\in S^{t+1}\setminus U^t}\dfrac{f(U^t\cup\{S_i\})-f(U^t)}{c_i}$.

上式第 2 个不等号根据 $f(U^t\cup\{S_i\})-f(U^t)\leqslant c_i\theta_{t+1}$ 与 $\sum_{S_i\in OPT\setminus U^t}c_i\leqslant L$ 综合可得. 令 $L_t=\sum_{r=1}^{t}c_{i_r}$, $L_0=0$, $(t=0,\cdots,t^*)$. 对于 $j=1,\cdots,\overline{L}=L_{t^*+1}$, 当 $j=L_{t-1}+1,\cdots,L_t$ 时, 定义 $\rho_j=\theta_t$. 由此可得

$$g(U^{t^*}\cup\{S_{t^*+1}\})=\sum_{r=1}^{t}c_{i_r}\theta_r=\sum_{j=1}^{\overline{L}}\rho_j,$$

$$g(U^t)=\sum_{r=1}^{t}c_{i_r}\theta_r=\sum_{j=1}^{L_t}\rho_j,\ (t=1,\cdots,t^*),$$

所以有

$$\min_{s=1,\cdots,\overline{L}}\left\{\sum_{j=1}^{s-1}\rho_j+L\rho_s\right\}=\min_{t=1,\cdots,t^*}\left\{\sum_{j=1}^{L_t}\rho_j+L\rho_{L_{t+1}}\right\}=\min_{t=1,\cdots,t^*}\left\{g(U^t)+L\theta_{t+1}\right\}.$$

文[5] 给出如下不等式

$$\frac{\sum_{i=1}^{P}\rho_i}{\min\limits_{t=1,\cdots,P}\left\{\sum_{i=1}^{t-1}\rho_i+D\rho_t\right\}}\geqslant 1-\left(1-\frac{1}{D}\right)^P>1-e^{-\frac{P}{D}}, \tag{5}$$

这里, D, P 为正整数, ρ_i, $(i=1,\cdots,P)$ 为非负实数且 $\rho_1>0$.

利用不等式(4) 和(5) 得

$$\frac{g(U^{t^*}\cup\{S_{t^*+1}\})}{g(OPT)}\geqslant\frac{\sum_{j=1}^{\overline{L}}\rho_j}{\min\limits_{s=1,\cdots,\overline{L}}\left\{\sum_{j=1}^{s-1}\rho_j+L\rho_s\right\}}\geqslant 1-e^{-\frac{\overline{L}}{L}}\geqslant 1-e^{-1}. \tag{6}$$

(下转第 13 页)

参 考 文 献:

[1] HUANG T Y, WENG C W. Pooling spaces and non-adaptive pooling designs [J]. Discrete Mathematics, 2004, 282: 163-169.

[2] MACULA A J. Error correcting nonadaptive group testing with d^e-disjunct matrices [J]. Disc Appl Math, 1997, 80: 217-222.

[3] YACHKOV A D, HWANG F K, MACULA A, et al. A construction of pooling designs with some happy surprises [J]. Journal of Computational Biology, 2005, 12: 1129-1136.

[4] HUANG T Y, WENG C W. A note on decoding of superimposed codes [J]. Journal of Combinatorial Optimization, 2003, 7: 381-384.

[5] WU W L, HUANG Y C, HUANG X, et al. On error tolerant DNA screening [J]. Discrete Applied Mathematics, 2006, 154: 1753-1758.

[6] PARK H, WU W L, LIU Z, et al. DNA screening, pooling designs and simplicial complex [J]. Journal of Combinatorial Optimization, 2003, 7: 389-394.

[7] NGO H Q, DU D Z. New constructions of non-adaptive and error tolerance pooling designs [J]. Discrete Mathematics, 2002, 243: 161-170.

[8] FU H L, HWANG F K. A novel use of t-packings to construct d-disjunct matrices [J]. Discrete Applied Mathematics, 2006, 154: 1759-1762.

[9] WAN Z X. Geometry of classical groups over finite fields [M]. Lund: Studentlitteratur/Bromley: Chartwell-Bratt, 1993.

(责任编辑: 李洪建)

(上接第 9 页)

设 H 为由上述算法求得的近似解, 由不等式(3) 和(6) 得

$$f(H) \geqslant f(U^{t^*}) = f(V) + g(U^{t^*}) = f(V) + g(U^{t^*} \cup \{S_{t^*+1}\}) - (g(U^{t^*} \cup \{S_{t^*+1}\})) - g(U^{t^*})) = f(V) + g(U^{t^*} \cup \{S_{t^*+1}\}) - (f(U^{t^*} \cup \{S_{t^*+1}\})) - f(U^{t^*})) \geqslant f(V) + (1 - e^{-1}) g(OPT) - f(V)/3 \geqslant (1 - e^{-1}) f(OPT).$$

参 考 文 献:

[1] SUDIPTO G, SAMIR K. Greedy strikes back: Improved facility location algorithms [J]. Journal of Algorithms, 1999, 31(6): 228-248.

[2] ODED B, DIMITRI B, RICHARD C L. Locating discretionary service facilities, Ⅱ: Maximizing market size, minimizing inconvenience[J]. Operations Research, 1995, 43(4): 623-632.

[3] BERMAN O, FOUSKA N, LARSON R C. Optimal location of discretionary service facilities[J]. Transportation Science, 1992, 26(3): 201-211.

[4] DORIT S.HOCHBAUM A. Approximation Algorithms for NP-hard problems[M]. Boston: PWS Publishing Company, 1996.

[5] FUJISHIGE S. Submodular functions and optimization[M]. Amsterdam: Elsevier Science, 2005.

[6] MAXIM S. A note on maximizing a submodular set function subject to knapsack constraint [J]. Operations Research Letters, 2004, 32(5): 41-43.

[7] CHANDRA C, AMIT K. Maximum coverage problem with group budget constraints and applications[J]. Lecture Notes in Computer Science, 2004, 12(2): 72-83.

[8] SAMIR K, ANNA M, JOSEPF N. Budgeted maximum coverage problem[J]. Information Processing Letters, 1999, 70(1): 39-45.

(责任编辑: 李洪建)