次模函数近似算法求最小弱顶点覆盖

涂建华 高昊宇 赖文华
(北京化工大学 理学院, 北京 100029)

摘 要: 求给定无向图的最小弱顶点覆盖是一个 NP 困难问题,只能通过研究此问题的近似算法来求解。本文从基本圈出发,定义了一个次模函数,利用次模函数理论来得到一个最小弱顶点覆盖问题的近似解,且近似度为 $1+\ln(d-1)$,其中 d 为图的顶点最大度。

关键词: 最小弱顶点覆盖; 次模函数; 近似算法; 近似度
中图分类号: O157

引 言

随着 Internet 应用的增长,网管系统越来越注重于应用服务的管理,网络流量监测需要更大的数据量和更高的数据采集频率。因此,为了优化网络性能,设置高效合理的流量监测系统就成了重要的研究课题[1-4]。已有的研究中一般将网络流量监测问题转化为求给定无向图中的最小弱顶点覆盖问题。但求解最小弱顶点覆盖是一个 NP 困难问题[1,5],只能通过近似算法求解。国内外学者对此问题已进行了深入研究[3-7],文献[3]提出了 3 个计算最小弱顶点覆盖的近似算法,并使用模拟数据加以比较。文献[4]在文献[3]的基础上,从图论角度提出了一个近似度为 $2(\ln d+1)$ 的弱化的贪婪算法,其中 d 为图的顶点最大度。文献[5]根据有关环路空间的性质,得到了一个近似度为 $\ln d+1$ 的贪婪算法。

本文从新的角度出发,应用次模函数理论得到最小弱顶点覆盖问题的一个近似解,且近似度为$1+\ln(d-1)$。

1 弱顶点覆盖及相关引理

定义 1(弱顶点覆盖)[4] 假定无向图 $G=(V,E)$ 满足对任意 $v\in V$ 有 $d(v)\geqslant 2$,称 $W\subseteq V$ 是图 G 的弱顶点覆盖集,当且仅当执行以下操作能使 E 中所有边可以被标记。

(1) 标记所有与 W 中顶点相关联的边;

(2) 若某个顶点 v 的 $d(v)-1$ 条相关联的边已被标记,则标记剩下的那条相关联的边;

(3) 重复第(2)步,直到不能再标记新的边为止。

若每个顶点 v 都有一个权重 $c(v)$,则最小弱顶点覆盖问题是寻找一个权重和最小的弱顶点覆盖集。

定义 2(基本圈)[8] 设 T 是图 G 的一个极大森林(边数达到最大),则对于每一条不在 T 上的边 e,$T\cup\{e\}$ 中都有唯一的一个圈与之相应,称这个圈为关于 T 的一个基本圈。

引理 1[4] 假定无向图 $G=(V,E)$ 满足对任意 $v\in V$ 有 $d(v)\geqslant 2$,则 $S\subseteq V$ 是图 G 的弱顶点覆盖集,当且仅当 $G'=(V',E')$ 是森林,其中 $V'=V-S$。

引理 2[5] 令 $G=(V,E)$,其中 $|V|$ 为顶点个数,$|E|$ 为边数,p 为其连通分支个数。令 $J=(V,T)$ 为图 G 的极大森林,对 $u\in\overline{T}$,记 μ^u 为 $T+\{u\}$ 上的唯一圈,当 u 历经 $\overline{T}$ 的所有边时,形成图 G 的圈空间的一个基 $\{\mu^u\}$,其维数为图 G 中基本圈的个数等于 $|E|-|V|+p$。

2 次模函数

定义 3(次模函数) 考虑有限集合 E 和定义在其幂集 2^E 的一个实函数 f,称 f 是次模函数当且仅当对于任意集合 $X\subset E, Y\subset E$ 满足

$$f(X)+f(Y)\geqslant f(X\cap Y)+f(X\cup Y) \qquad (1)$$

记 $\Delta_D f(Z)=f(Z\cup D)-f(Z)$,其中 Z、D 为 E 的任意子集。此时式(1)可表示为

收稿日期: 2010-05-25
基金项目: 中央高校基本科研业务费专项资金(QN0913)
第一作者: 男,1981 年生,讲师,博士
E-mail: tjhyuyong@yahoo.com.cn

$$\Delta_D f(X\cap Y)\geqslant \Delta_D f(Y)$$

其中 $D=X\setminus Y$，当 D 为集合 $\{x\}$ 时，把 $\Delta_{\{x\}}f(Z)$ 简单记为 $\Delta_x f(Z)$。

这里举一个次模函数的例子。对 E 的任意两个子集 X、Y，定义 f 为集合的元素个数，即集合的势。则有 $|X|+|Y|=|X\cap Y|+|X\cup Y|$，这说明集合的势是一个次模函数。显然该例中等号恒成立，称这样的 f 为模函数。

对于函数 $f:2^E\to R$，当 $X\subset Y$ 时，有 $f(X)\leqslant f(Y)$，则称 f 是单调递增的。

考虑定义在 2^E 上的次模函数 f，记 $\Omega_f=\{Z|\Delta_x f(Z)=0,\forall x\in E\}$。下面对于单调递增的次模函数，给出一个常用的贪婪算法。

令 c 是定义在 E 上的非负费用函数，记 $c(X)=\sum\limits_{x\in X}c(x)$，考虑最小化问题

$$\min\quad c(X)$$
$$\text{s.t.}\quad X\in\Omega_f$$

这个最小化问题是许多问题的一般形式，可由贪婪算法求解。

贪婪算法 1

(1) 输入次模函数 f 和费用函数 c。

(2) $X\leftarrow\varnothing$。

(3) 若存在 $x\in E$ 使得 $\Delta_x f(X)>0$，则转步骤(4)；否则转步骤(5)。

(4) 选择一点 x 使得 $\dfrac{\Delta_x f(X)}{c(x)}$ 最大，转步骤(2)。

(5) $X\leftarrow X\cup\{x\}$。

(6) 输出 X。

定理 1[9]　f 为单调递增的次模整数函数，针对上述最小化问题设计的贪婪算法的近似比为 $H(\gamma)$，其中 $\gamma=\max_{x\in E}\Delta_x f(\phi)$，$H(\gamma)=\sum\limits_{i=1}^{\gamma}\dfrac{1}{i}$ 为调和函数。

3　利用次模函数理论求解弱顶点覆盖

考虑无向图 $G=(V,E)$，其中 V 为图 G 的顶点集，E 为图 G 的边集，对任意 $S\subseteq V$，令 D_S 为点集 S 所关联的边数，$|S|$ 为点集 S 中点的个数，k 为图 G 的连通分支个数，k_S 为图 $G-S$ 的连通分支个数。定义 $f:2^V\to R$

$$f(S)=D_S-|S|+k-k_S,\ \forall S\subseteq V。$$

引理 3　f 是单调递增函数，即对

$\forall S_1,S_2\subseteq V$，若 $S_1\subseteq S_2$，则 $f(S_1)\leqslant f(S_2)$。

证明：设顶点 $v\in S_2\setminus S_1$，只需证 $f(S_1)\leqslant f(S_1\cup\{v\})$ 成立，对顶点 v 分情况讨论。

(1) 若 v 的所有邻点都在 S_1 中，即 $N_G(v)\subseteq S_1$，$N_G(v)$ 为 v 在图 G 中的邻点集。

此时有 $D_{S_1}=D_{S_1\cup\{v\}}$，$|S_1\cup\{v\}|=|S_1|+1$，因为 $N_G(v)\subseteq S_1$，所以在图 $G-S_1$ 中，点 v 为一个孤立点，故 $k_{S_1\cup\{v\}}=k_{S_1}-1$，则

$$f(S_1\cup\{v\})=D_{S_1\cup\{v\}}-|S_1\cup\{v\}|+k-k_{S_1\cup\{v\}}=D_{S_1}-(|S_1|+1)+k-(k_{S_1}-1)=f(S_1)$$

(2) 若 v 有 i 个邻点不在 S_1 中，即 $|N_G(v)\setminus S_1|=i$；此时有 $D_{S_1}=D_{S_1\cup\{v\}}+i$，$|S_1\cup\{v\}|=|S_1|+1$。在 $G-S_1$ 中，当把 v 从 $G-S_1$ 中删除时，v 所在的连通分支最多分成了 i 个连通分支。故有 $k_{S_1\cup\{v\}}\leqslant k_{S_1}+i-1$，所以

$$f(S_1\cup\{v\})=D_{S_1\cup\{v\}}-|S_1\cup\{v\}|+k-k_{S_1\cup\{v\}}\geqslant D_{S_1\cup\{v\}}+i-(|S_1|+1)+k-(k_{S_1}+i-1)=f(S_1)$$

综上，f 是单调递增函数。

因为 f 是单调递增函数，所以对于 $\forall S\subseteq V$，$f(S)\leqslant f(V)=|E|-|V|+k$。

引理 4[9]　设 $f:2^E\to R$，f 为次模函数当且仅当对于集合 $X\subseteq Y$，及任意元素 $y\notin Y$，有 $\Delta_y f(X)\geqslant\Delta_y f(Y)$ 成立。

引理 5　f 为次模函数。

证明：由引理 4，只需要证明若 $S_1\subseteq S_2\subseteq E$，$\forall v\notin S_2$，$\Delta_v f(S_1)\geqslant\Delta_v f(S_2)$。由定义知，

$$\Delta_v f(S)=f(S\cup\{v\})-f(S)=D_{S\cup\{v\}}-D_S-1-k_{S\cup\{v\}}+k_S。$$

对 v 分类讨论。

(1) v 的所有邻点都在 S_2 中，即 $N_G(v)\subseteq S_2$；此时有 $D_{S_2}=D_{S_2\cup\{v\}}$，$|S_2\cup\{v\}|=|S_2|+1$，因为 $N_G(v)\subseteq S_2$，所以在图 $G-S_2$ 中，点 v 为一个孤立点，故 $k_{S_2\cup\{v\}}=k_{S_2}-1$，则

$$f(S_2\cup\{v\})=D_{S_2\cup\{v\}}-|S_2\cup\{v\}|+k-k_{S_2\cup\{v\}}=D_{S_2}-(|S_2|+1)+k-(k_{S_2}-1)=f(S_2)$$

即有 $\Delta_v f(S_2)=f(S_2\cup\{v\})-f(S_2)=0$。又由 $f(S)$ 的单调递增性有 $f(S_1\cup\{v\})\geqslant f(S_1)$，则 $\forall v\notin S_2$，$\Delta_v f(S_1)\geqslant\Delta_v f(S_2)=0$；

(2) v 中有 j 个邻点不在 S_1 中，有 i 个不在 S_2 中，

且 $j \geqslant i \geqslant 1$,则

$|N_G(v) \backslash S_1| = j, |N_G(v) \backslash S_2| = i$。

此时 $D_{(S_1 \cup \{v\})} - D_{S_1} = j, D_{(S_2 \cup \{v\})} - D_{S_2} = i, |S_1 \cup \{v\}| = |S_1| + 1, |S_2 \cup \{v\}| = |S_2| + 1$。$v$ 中有 $j-i$ 个邻点不在 S_1 中但在 S_2 中。当把 v 从 $G-S_2$ 中删除时,连通分支的个数增加 $k_{S_2 \cup \{v\}} - k_{S_2}$。因为 $G-S_2$ 为 $G-S_1$ 的子图,当把 v 从 $G-S_1$ 中删除时,连通分支增加的个数有 $k_{S_1 \cup \{v\}} - k_{S_1} \leqslant k_{S_2 \cup \{v\}} - k_{S_2} + j - i$。因此由

$$\Delta_v f(S_1) = f(S_1 \cup \{v\}) - f(S_1) = D_{S_1 \cup \{v\}} - D_{S_1} - 1 - k_{S_1 \cup \{v\}} + k_{S_1}$$

和 $D_{S_1 \cup \{v\}} - D_{S_1} = D_{S_2 \cup \{v\}} - D_{S_2} + j - i$ 可得 $\Delta_v f(S_1) \geqslant \Delta_v f(S_2)$。

综上,f 为次模函数。

引理 6 若 $\Omega_f = \{S | \Delta_x f(S) = 0, S \subseteq V\}$,则 $S \in \Omega_f \Leftrightarrow S$ 为 G 的弱顶点覆盖。

证明:

先证充分性。

若 S 为弱顶点覆盖,则由引理 1 知此时 $G-S$ 为森林即无基本圈,即 $|E| - D_S - (|V| - |S|) + k_S = 0$,所以 $D_S - |S| + k - k_S = |E| - |V| + k$,由 $f(S)$ 定义知 $f(S) = |E| - |V| + k$ 达到了最大值,所以 $S \in \Omega_f$。

再证必要性。

若 $S \in \Omega_f$,即 $\Delta_x f(S) = 0$,假设 S 不是 G 的弱顶点覆盖,那么 $G-S$ 中仍有圈 C,v 为圈 C 上的任意一点。设 v 的所有邻点中有 i 个不在 S 内($i \geqslant 2$)。

因为在图 $G-S$ 中,$v \in C$,所以当把点 v 从 $G-S$ 中删除时,v 所在的连通分支最多分割成 $i-1$ 个,即有 $k_{S \cup \{v\}} \leqslant k_S + i - 2$,故,

$$\begin{aligned} f(S \cup \{v\}) &= D_{S \cup \{v\}} - |S \cup \{v\}| + k - k_{S \cup \{v\}} \geqslant \\ &D_s + i - (|S| + 1) + k - (k_s + i - 2) \geqslant \\ &D_s - |S| + k - k_S + 1 \geqslant f(S) \end{aligned}$$

这与 $S \in \Omega_f$ 矛盾,所以假设不成立,即图 $G-S$ 中不含有圈,由引理 1,S 为弱顶点覆盖。

定理 2 由次模函数理论导出的贪婪算法近似比为 $H(d-1)$ 且 $H(d-1) \leqslant 1 + \ln(d-1)$。

证明:在弱顶点覆盖问题中,对任意 $v \in V$ 有 $d(v) \geqslant 2$,故 $\forall v \in V, k \leqslant k_v$。由定理 1

$$\gamma = \max_{v \in V} \Delta_v f(\varphi) = \max\ (D_v - 1 + k - k_v) \leqslant D_v - 1 = d - 1$$

其中 d 为图中顶点的最大度。所以该贪婪算法的近似度为 $H(d-1) = \sum_{i=1}^{d-1} \frac{1}{i} \leqslant 1 + \ln(d-1)$。

4 结论

建立了最小弱顶点覆盖问题的相应次模函数 $f(S) = D_S - |S| + k - k_S$,并有效地应用次模函数相关理论设计近似算法,求解最小弱顶点覆盖问题,且得到了近似度为 $1 + \ln(d-1)$,其中 d 为图的顶点最大度。

参考文献:

[1] 刘湘辉, 殷建平, 卢锡城, 等. 基于弱顶点覆盖的网络链路使用带宽检测模型[J]. 软件学报, 2004, 15(4): 545-549.
Liu X H, Yin J P, Lu X C, et al. A monitoring model for link bandwidth usage of network based on weak vertex cover[J]. Journal of Software, 2004, 15(4): 545-549. (in Chinese)

[2] 蔡志平, 刘芳, 赵文涛, 等. 网络测量部署模型及其优化算法[J]. 软件学报, 2008, 19(2): 419-431.
Cai Z P, Liu F, Zhao W T, et al. Deploying models and optimization algorithms of network measurement [J]. Journal of Software, 2008, 19(2): 419-431. (in Chinese)

[3] Breibart Y, Chan C Y, Carofalakis M, et al. Efficiently monitoring bandwidth and latency in IP network [C] // Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies Proceedings IEEE, April 22 – 26, 2001, Anchorage, AK, USA. 2001, 2: 933–942.

[4] 刘湘辉, 殷建平, 唐乐乐, 等. 网络流量的有效测量方法分析[J]. 软件学报, 2003, 14(2): 300-304.
Liu X H, Yin J P, Tang L L, et al. Analysis of Efficient Monitoring Method for the Network Flow[J]. Journal of Software, 2003, 14(2): 300-304. (in Chinese)

[5] Zhang Y, Zhu H. Appromation algorithm for weighted week vertex cover[J]. Journal of Computer Science and Technology, 2004, 19(6): 782-786.

[6] 彭震宇. 最大独立集和最小弱顶点覆盖问题求解及其应用研究[D]. 无锡: 江南大学, 2008.
Peng Z Y. Research of maximum independent set and minimum weak vertex cover problem solving and application[D]. Wuxi: Jiangnan University, 2008. (in Chinese)

[7] 石恒华, 何泾沙, 许鑫. 基于三元组信息的网络流量监测点选取算法[J]. 北京邮电大学学报, 2009, 32

（增 1）：73-76.
Shi H H, He J S, Xu X. A network traffic monitor-node selection algorithm based on triple tag[J]. Journal of Beijing University of Posts and Telecommunications, 2009, 32(z1): 73-76. (in Chinese)

[8] 马登举，刘凯峰. 两个不交图的联图的最小圈基长度[J]. 应用数学学报，2008，31(5)：845-851.
Ma D J, Liu K F. The Length of the Minimum Cycle Bases of Join of Two Disjoint Graphs[J]. Acta Mathematicae Appicatae Sinica, 2008, 31(5): 845-851. (in Chinese)

[9] 李学良，涂建华. 次模函数近似算法求最小颜色生成树[J]. 新疆大学学报：自然科学版，2008，25(4)：391-394.(英文)
Li X L, Tu J H. Submodular Potential Function for the Minimum Color Spanning Tree Problem of Edge-colored Graphs[J]. Journal of Xinjiang University: Natural Science Edition, 2008, 25(4): 391-394.

Submodular potential function for the minimum weak vertex cover problem

TU JianHua　GAO HaoYu　LAI WenHua
(School of Science, Beijing University of Chemical Technology, Beijing 100029, China)

Abstract: The minimum weak vertex cover problem is non-deterministic polynomial-time hard (NP-hard), and thus we can only produce an approximate solution to this problem. Here we start from a fundamental cycle to define a submodular potential function. Then using the theory of submodular potential functions, we give an approximation algorithm to solve the problem. The approximation factor of the algorithm is $1 + \ln(d-1)$, where d is the maximum degree of the vertices.

Key words: minimum weak vertex cover; submodular potential function; approximation algorithm; approximation factor