第10卷 第20期 2010年7月
1671-1815(2010)20-4891-05

科 学 技 术 与 工 程
Science Technology and Engineering

Vol. 10 No. 20 July 2010

论 文

数 学

求一般图的最小顶点覆盖集问题的混合贪婪算法

吕健康 张国基
(华南理工大学理学院,广州 510000)

摘 要 现有的求一般图的最小顶点覆盖集近似算法或者近似比较高,或者为降低复杂度限制了图的规模,或者算法搜索过程中盲目性大。根据顶点的度特点及贪婪法的思想,提出了邻接度数、覆盖边等主要概念,并在此概念的基础上设计了混合贪婪算法。该算法设计思路清晰,容易理解,易于编程实现,且在最坏情况下的时间复杂度为 $O(|V|^2)$,执行效果较好,性能近似比不大于4/3,接近已知的可能的近似比下界1.166 6,低于2005年认为最低的近似比1.361,是图的最小顶点覆盖问题算法的一个较好的补充。

关键词 最小顶点覆盖集 复杂度分析 混合贪婪算法
中图法分类号 O157.5; **文献标志码** A

最小顶点覆盖集(the Minimum Vertex Cover Problem,简称Min-VC)问题是一个NP完全问题[1],然而Min-VC问题在工程中应用非常广泛,许多工程中问题都属于或可以转化为图的Min-VC问题[2,3]。为了求解图的Min-VC问题,国内外学者提出了基于布尔运算的方法[4]、DNA算法[5]、基于神经网络的算法[6,7]、混合遗传算法[8]等,这些算法都是时间复杂度关于图的规模或者迭代次数呈指数型增长,当图的规模或者迭代次数越来越大时,会消耗大量的时间,并且空间复杂度也越来越大;或者以某种概率求得最优解。为此,学者们提出了各种近似算法[9—12],但这些算法或者时间复杂度较高,或者限制了问题的规模,或者性能近似比差,或者算法搜索过程中有一定的盲目性,或者执行效果差,从而影响了算法的应用。文献[3]对参数化点覆盖问题进行了研究。

本文根据顶点的度特点及贪婪法的思想,提出了邻接度数、覆盖边等概念,并在此概念的基础上设计了混合贪婪算法(the Mixed Greedy Algorithm,简称MGA)。该算法在最坏情况下时间复杂度为 $O(|V|^2)$,空间复杂度 $O(|V|+|E|)$,近似比为4/3,接近于目前已知的可能的近似比下界1.166 6[4]。低于2005年被认为是最低的近似比 $10\sqrt{5}-21 \approx 1.361$[5,16]。且执行效果较好。

1 相关概念

对于无向图 $G=(V,E)$,其中 V 是顶点的集合,E 是边的集合,设 $v \in V$,$e \in E$,若 v 与 e 关联,则称点 v 覆盖了边 e,设顶点集 $V_1 \subseteq V$,若 E 中的每条边都与 V_1 中的某点关联,则称 V_1 是图 G 的一个顶点覆盖集,若对图 G 的任意顶点覆盖集 V_2 均有 $|V_1| \leqslant |V_2|$ 则称 V_1 是图 G 的最小顶点覆盖集,记作 $\alpha(G)$。通常,一个图的最小顶点集不是唯一的。

2010年4月21日收到
第一作者简介:吕健康(1984—),男,华南理工大学硕士研究生,研究方向:计算智能、图像处理。

称最小覆盖集中的顶点为最小覆盖点。注意，孤立顶点不是最小覆盖点。

称图 G 中所有与点 v 相邻的点的集合为 v 的邻域，记作 $N_G(v)$ 。设图 G 的任意一个顶点的子集 V_1，称 $N_G(V_1) = \{ v \mid v \in V$，且 v 与 V_1 中某点相邻接$\}$为 V_1 的领域。定义顶点 v 的边领域为 $E(v) = \{e \mid e \in E$ 且 e 与 v 关联$\}$，顶点集 V_1 的边领域 $E(V_1) = \{ e \mid e \in E$ 且 e 与 V_1 中某点关联$\}$。

设图 $G_i = (V_i, E_i)$ ，$i = 1,2$ ，若 $V_1 \subseteq V_2$ ，$E_1 \subseteq E_2$ ，则称 G_1 为 G_2 的子图，记作 $G_1 \subseteq G_2$ 。

若非空集 $F \subseteq E(G)$ ，则 $G - F$ 表示从 G 中删除 F 中所有边得到的子图，并将 $G - \{e\}$简记为 $G - e$ ，若非空子集 $S \subseteq V(G)$ ，则 $G - S$ 表示从 G 中删除 S 中所有顶点及与 S 中顶点关联的所有边后得到的子图，并将 $G - \{ v \}$简记为 $G - v$ 。

图 G 中顶点 v 的度定义为以该顶点为一个端点的边的数目，简单地说就是该顶点的边的数目，记为 $D(v)$ 。

令问题$\prod$是一个最小化问题，I 是问题的一个实例，A 是解问题$\prod$的一个近似算法，$A(I)$是用算法 A 对问题$\prod$的实例 I 求解时得到的近似值，OPTA 是解问题$\prod$的最优算法，$OPTA(I)$是算法 OPTA 对问题$\prod$的实例 I 求解时得到的准确值，在本文里，指的是最小覆盖集的顶点数，则算法 A 的性能近似比为：

$$\rho(I) = \frac{A(I)}{OPTA(I)} \tag{1}$$

2 算法的描述及其性能分析

2.1 算法的主要概念

定义 1 对图 $G = (V,E)$ ，设 v 是图 G 的任意一个顶点，v 的度数和与 v 关联的所有顶点的度数之和称为顶点 v 的邻接度数，记为 $L(v)$ ，计算公式如式(2)。

$$L(v) = \sum_{v' \in \mathrm{N}_G(v)} D(v') + D(v) \tag{2}$$

定义 2 对图 $G = (V,E)$ ，设顶点集 $V_1 \subseteq V$ 是图 G 的一个顶点覆盖集，则任意的边 $e \in E$ 就称为顶点覆盖集 V_1 的一个覆盖边。

2.2 基于图的顶点邻接度数的 MGA 算法

本文根据顶点的度特点及贪婪法的思想，设计了 MGA 算法。该算法分为两个阶段，第一阶段选取顶点的邻接度数较大的顶点，以使得最少数目的图的顶点并入覆盖集里为原则，即使得最多数目的图的边成为覆盖边；第二阶段是经过第一阶段的搜索，得到图 G 的子图，计算子图的顶点邻接度数，按照同样的方法，最终找到最小顶点覆盖集。

为了便于搜索，我们将图的算法使用访问标记，不同的标记决定了下一步的搜索与选择，用$f_i(0 \leqslant i \leqslant n-1)$ 表示顶点i 的搜索标记，不同值表示的意义为：f_i 为 -1，表示没有访问过的顶点；f_i 为 0，表示顶点已访问过。

为了方便起见，图的顶点用数字编号，令图的顶点集合为 $V = \{0,1,\cdots,n-1\}$。本文用图的邻接表来表示图中各顶点及其关联边之间的关系。

此外，根据算法的需要定义下面的数组和三个变量：$c[n]$ 表示存储图 G 的顶点覆盖集；V_count 是记录访问顶点的个数；V'_count 是记录删除顶点的个数；E_count 是记录处理图 G 的边数。

现算法描述步骤如下：

(1)初始设置 $f_i = -1$，$(0 \leqslant i \leqslant n-1)$，$V\_count = 0$，$E\_count = 0$，$V'\_count = 0$；

(2)利用邻接表存储图 G 的顶点特点，计算每个顶点的度数，若某顶点的度数为 0，删除此顶点；

(3)根据公式(2)，计算每个顶点的邻接度数；

(4)选取邻接度数最大的顶点 i，登记到顶点覆盖集 C 中；

(5)处理顶点 i：删除顶点 i，则 $V'\_count{+}{+}$ ，$|V| = |V| - V'\_count$；并且删除所有与顶点 i 关联的边，每删除一边，则 $E\_count{+}{+}$；并且还把所有与顶点 i 邻接的顶点 j 标记为访问过，即$f_j = 0$，每访问一个顶点，则 $V\_count{+}{+}$；

(6)若 $V\_count \geqslant |V|$，转步骤(7)；否则，算法继续搜索，选取邻接度数最大的顶点 k，若顶点 k 已访问，则不作处理，若没有访问过，转步骤(5)；

(7)若 $E\_count \geqslant |E|$ ，算法结束，否则，处理图

G 的子图，令把子图的所有顶点都标记为未访问过，即 $f_i = -1$，，且 $V\_count = 0$，$V'\_count = 0$，转步骤（2）。

2.3 算法时间复杂度分析

分析算法的步骤可知，第一步顶点的访问标记初始化，需花费 $\Theta(|V|)$ 时间，第 2、3 步计算顶点的度数及邻接度数根据邻接表的存储特点计算得到的，需 $\Theta(2|E|)$ 时间，算法第 5 ~ 7 步执行时间由两部分组成：第一部分是选择最大的邻接度数，总花费 $\Theta(n-1+\cdots+2+1) = \Theta(|V|^2)$；第二部分顶点访问标记和删除邻接边，总花费 $\Theta(|E|+|V|)$ 时间。由此算法总的花费时间是 $\Theta(2|E|+2|V|+|V|^2)$，当 $|E| = O(|V|^2)$ 时，算法的花费总时间为 $O(|V|^2)$。MGA 算法的时间复杂度为多项式时间 $O(n^2)$，n 为图的顶点个数，当随图的规模增大时，运行时间在最坏的情况下才是多项式 n^2 的增加。

2.4 算法的空间复杂度分析

MGA 算法除了存放作为输入用的邻接表需要 $\Theta(|E|)$ 的空间外，还用于存放顶点覆盖集所需的工作单元，需要 $\Theta(|V|)$ 的工作空间。故 MGA 算法的空间复杂为 $\Theta(|E|+|V|)$。

3 算法性能分析

对于无向图 $G = (V,E)$，设 $C \subset V$ 使 G 的一个最小顶点覆盖集，则存在 C 的一个划分 $\pi = \{C_1, C_2 \cdots, C_n\}$。容易验证，若 G 的顶点子集 C'_i 覆盖了 C_i 覆盖的所有边，且 $|C'_i| \leqslant r_i |C_i|$，$i = 1,2,\cdots,n$，则 $C' = \cup C'_i$ 也是图 G 的一个顶点覆盖，且 $|C'|/|C| \leqslant \max r_i$。称 r_i 为算法所取得的最小覆盖集 C_i 所覆盖的那些边的局部近似比，简称算法的局部近似比，或局部近似比 [2]。事实上，一旦算法选取了 C'_i，则必相应存在最小顶点覆盖集的一个子集 C_i 来覆盖边集 $E(C'_i)$，所以计算算法的局部近似比从而推断算法的近似比是可行的 [2]。我们根据算法的设计，把图的顶点分为两类：一种是顶点度数 $D(v) \geqslant 3$，另一种是 $D(v) \leqslant 2$。

定理 1 [2] 算法处理顶点度数 $D(v) \geqslant 3$ 的局部近似比不大于 4/3。

证明 设 $D(v) \geqslant 3$，则 $|N(v)| \geqslant 3$。任意一最小顶点覆盖集 C 均覆盖 $E(v)$ 上的所有边。若 $v \in C$，则算法的局部近似比为 1。否则，在最坏的情况下，$N(v)$ 完全包含于 C 分布在 $N(v)$ 上的子集，此时算法选取了集合 $V = \{v\} \cup N(v)$ 来覆盖 $E(v)$ 上的边，此时，局部近似比为 $\frac{|N(v)|+1}{|N(v)|} \leqslant 4/3$。

定理 2 算法处理顶点度数 $D(v) \leqslant 2$ 的局部近似比 1。

证明 所有度数 $D(v) \leqslant 2$ 的顶点连通构成的子图，或者是链状 $A_1 - A_2 - \cdots - A_r$，或者是环状。若是链状，由此算法得到的顶点覆盖是从 A_2，$A_4 \cdots$ 每隔一个选取得到的，近似比是 1。若是环状，由算法得到的顶点覆盖由也是从环的一个顶点开始每隔一个选取得到，近似比也是 1。

于是，由定理 1、2 可以判断：MGA 算法的近似比不大于 4/3。可见，MGA 算法的性能近似比优于文献 [0] 的顶点覆盖问题的贪心算法的性能近似比 $H(d) = \sum 1/j$，$(j = 1,2,\cdots,d)$，d 为图中最大的顶点度数。

4 算法实例

在本文里以图 1 实例，MGA 算法和文献 [0] 和文献 [2] 进行试验比较。图 2 表示 MGA 算法的处理过程。图 2a 表示开始搜索图 G，把图 G 的所有顶点标记为未访问，并计算顶点的度及邻接度数；图 2b 根据邻接度数的大小，依次对图 G 的顶点进行处理，得到图 G 的子图及最小覆盖点 h、a、c、e、m；图 2c 进入第二阶段的搜索，搜索图 G 的子图把所有的变量复原，计算子图的顶点的度数和邻接度数，并把顶点标记为未访问过；又得到最小覆盖点 k、i，最后得到图 G 的一个最小覆盖顶点集 C = {h、a、c、e、m、k、i}，即为所有求。（见图）。

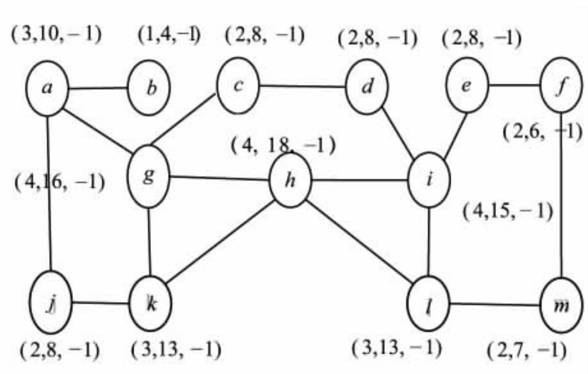
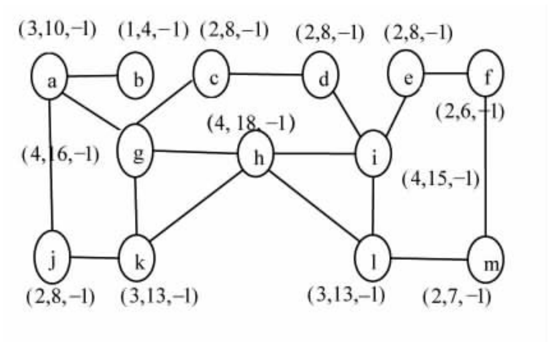


图 1 原始图



a 第一阶段图 G 的顶点的度、邻接度数和访问情况

b 第一阶段算法的访问后的结果

c 第二阶段图G 顶点的度、邻接度数和访问情况

图 2 算法搜索实例

而由文献［0］中的算法得到图 1 的最小顶点覆盖集为{g,i,a，f，h,j,c,m}，文献［2］中的算法得到的结果为{a,g,h,I,e,f,l,h,k}，可见，这两个文献得到的结果都劣于 MGA 算法。

文献［0］中的算法是从每次选择边转去选择度数最大的顶点，而在很多图中有很多顶点的度数是一样的大，这样就造成了有很大的盲目性选择，且文献［2］中的算法也是在判断一个顶点所属的局部结构类型的时候有多个链要选，也造成了一定的盲目性选择；还有就是算法每步选择顶点时，如果顶点的度数大于或等于 3 时就取作最小顶点覆盖集中，而在很多图中顶点的度数大于或等于 3，这样也造成了执行效果差的原因。

MGA 算法是每步都是选择邻接度数大的顶点来处理的，每访问完所有顶点后，重新计算邻接度数，这样在一定程度上克服了盲目性的选择。

5 结 论

1）本文根据顶点的度特点及贪婪的思想，提出了求解一般图的最小顶点覆盖问题的混合贪婪算法。

2）相对于同类算法，本文提出的算法对图的规模没有限制且在最坏情况下的时间复杂度为 $O(|V|^2)$，性能近似比不大于 4/3，接近于目前已知的可能的近似比下界 1.166 6 [2]。低于 2005 年被认为是最低的近似比 $10\sqrt{5}-21\approx 1.361$ [3,14]。且执行效果较好。

3）该算法设计思路清晰，容易理解，易于编程实现，搜索过程中有较好的选择性，且是图的最小顶点覆盖问题的方法中一个较好的算法。

参 考 文 献

1 郑宗汉，郑晓明．算法设计与分析．北京：清华大学出版社，2005：297—298

2 蔡志平，殷建平，刘湘辉．延迟约束的分布式演化网络监测模型．软件学报，2006；17(1)：117—123

3 Cai Zhiping，Yin Jianping，Lu Xianghui，*et al*. Efficently monitoring link bandwidth in IP net-works. Proc IEEE GLOBECOM，2005：St Louis，USA：［sn］，2005

4 刘永才，张 卫. 布尔方法论. 上海：上海科学技术文献出版社，1993：401—406
5 Chen Zhiyun，Qu Huiqin，Lu Mingning，*et al*. Apr-obabilistic parametetized algorithm for ve-rtex cover in stickermodel，parallel and distributed processing symposium. Proce-ssdings 18th Intemational New York：IEEE Xplore，2004：1859—1862
6 Yuan S Y，Kuo S Y. A new technique for optiniz-ation problems in graph theory. IEEE Transa-ctions on computers，1998；47（2）：190—196
7 余荣道. 基于人工神经网络的图的顶点覆盖问题的算法研究. 武汉：华中科技大学，2003：31—36
8 王 成，周育人，涂卫平. 一种求解顶点覆盖问题的混合遗传算法. 计算机工程与应用，2007；43（14）：27—41
9 肖鸣宇，陈建二，韩旭里. 低度图的点覆盖和独立集问题的下界改进. 计算机学报，2005；28（2）：153—160
10 姚明灼. 顶点覆盖问题的贪心算法的设计和分析. 福州大学学报（自然科学版），2001；29（1）：8—11
11 宁爱兵，马 良，熊小花. 最小顶点覆盖快速将阶算法. 小型微型计算机系统，2008；29（7）：1282—1285
12 闫兴篡，殷建平，蔡志平，等. 求图的最小顶点覆盖集的一个近似算法. 哈尔滨工业大学学报，2008；40（7）：1131—1135
13 Chen J. Parameterized computation and compl-exity：a new approch dealing with NP-hardness Journal of Computer Science and Technology，2005；20：18—37
14 Hastad J. Some optimal inapproximability results. Proc 29th Ann ACM Symp on Theory of Comp. New York：ACM，1997：1—10
15 Chen J，Ia K J. On approximating minimum vertex cover for graphs with prefect matching. Theoretical Computer Science，2005；337：305—318
16 Dnur I，Safra S. The importance of being b-aised on Theory of Computing. Proc 34th Annu ACM Symp. New York：ACM，2002：33—42

A New Mixed Greedy Algorithm for Solving Minimum Vertex Cover Set Problems

LÜ Jian-kang，ZHANG Guo-ji
（School of Science，South China University of Technology，Guangzhou 510000，P. R. China）

【Abstract】 Existing approximate algorithms for searching the minimum vertex cover set of a given graph either have a higher ratio bound or constrain the scale of the graph intending to reduce the computational complexity or shows big blindness for the algorithm searching . Based on the characteristics of the vertex degree and the idea of the greedy algorithm，the two primary concepts of abutting degree and vertex cover border are proposed. According to these concepts，a new mixed greedy algorithm is proposed. This proposed algorithm shows a higher efficienct，more understandable and easier for computer implementation. Theoretically，it has a computational time of $O(|V|^2)$ and a ratio bound of 4/3. This ratio bound is close to the known potential ratio bound of 1.166 6，and better than the result of 1.361 in 2005. The algorithm is a valuable alternative to find the minimum vertex cover set of a given graph.

【Key words】 minimum vertex cover problem complexity analysis mixed greedy algorithm