文章编号:1001 - 9081(2008)12 - 3072 - 02

最大团问题的改进遗传算法求解

吴冬晖,马　良
(上海理工大学 管理学院,上海 200093)
(hui_knight@ msn. com)

摘　要:最大团问题是组合优化中经典的 NP 完全问题,该问题的枚举算法只适用于求解中小规模的图。提出了基于遗传算法的最大团问题求解算法,引入概率模型指导变异产生新的个体,并结合启发式局部算法搜索最大团。经算例测试,获得了较好的效果。

关键词:最大团问题; 优化; 遗传算法

中图分类号: TP301.6　　**文献标志码**:A

Improved genetic algorithm for maximum clique problem

WU Dong-hui, MA Liang
(*Business School, University of Shanghai for Science and Technology, Shanghai 200093, China*)

Abstract: Maximum Clique Problem (MCP) is a classic NP Complete problem in combinatorial optimization. Its enumeration algorithm is only suitable for graphs with small scale. A new method based on genetic algorithm for MCP was proposed. Probability model was introduced to guide mutation when generating offspring, and heuristic local search was combined to find clique. The experimental results show that its performance is satisfying.

Key words: Max Clique Problem (MCP); optimization; genetic algorithm

1　最大团问题

最大团问题(Maximum Clique Problem, MCP)是图论中的经典组合优化问题[1],也是一类 NP 完全问题。人工智能、聚类分析、信号传输、编码理论、移动计算、故障诊断等领域中许多问题都可以归结为最大团问题[2]。最大团、顶点覆盖、最大独立集是三个等价的 NP 完全问题[3],在多项式时间算法(P)问题与非多项式时间算法(NP)问题的研究中发现,NP 完全问题的复杂性与整个类的复杂性相关,这些问题中的任一问题如果存在多项式时间算法,那么所有 NP 问题都是多项式时间可解的。

1.1　基本概念

设 $G = (V,E)$ 为简单图,$V' \subseteq V$,V' 称为 G 的团是指 V' 中任何两个顶点均在 G 中相邻。团的顶点个数称为团数,称顶点数为 K 的团为 K- 团。团分为极大团和最大团,如果一个团不被其他任一团所包含,即它不是其他任一团的真子集,则称该团为图 G 的极大团。极大团中顶点数最多的团称为图 G 的最大团。

1.2　数学模型

最大团问题作为整数规划问题可以用一个二次 0-1 分配模型来描述[4]:

$$\max f(\boldsymbol{x}) = \sum_{i=1}^{n} x_i$$

$$\text{s.t.} \quad \begin{cases} x_i + x_j \leqslant 1, \forall (v_i, v_j) \in \overline{E} \\ x_i \in \{0,1\},\ i = 1,2,\cdots,n \end{cases} \tag{1}$$

若 x^* 是上式的最优解,则图 G 的一个最大团为 C,且 $|C| = f(x^*)$,若 $x_i = 1$,则 $v_i \in C$;若 $x_i = 0$,则 $v_i \notin C$。

MCP 同样等价于全局二次 0-1 问题[5]:

$$\min f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x}$$

$$\text{s.t.} \quad \boldsymbol{x} \in \{0,1\}^n$$

其中,$\boldsymbol{A} = \boldsymbol{A}_{\bar{G}} - \boldsymbol{I}$,$\boldsymbol{A}_{\bar{G}}$ 表示图 G 补图的邻接矩阵。若 x^* 是其最优解,则图 G 存在最大团 C,且 $|C| = -f(x^*)$。

2　遗传算法求解最大团

2.1　适应度函数

给定无向图 $G = (V,E)$,顶点集 V 的子集 $U \subset V$,将 U 编码为:

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_n) \in \{0,1\}^n \tag{2}$$

其中,若顶点 i 在 U 中,则 $x_i = 1$,否则 $x_i = 0$,因此搜索空间即为 $\Omega = \{0,1\}^n$。

最大团问题的适应度函数可以简单地使用团数来表示:

$$f(x) = \begin{cases} \sum_{i=1}^{n} x_i, & U \text{为团} \\ 0, & \text{其他} \end{cases} \tag{3}$$

2.2　搜索空间分解

对于任意阶数为 n 的图,其搜索空间 $\Omega = \{0,1\}^n$,若将 Ω 分解为 $n+1$ 个子搜索空间,则有:

$$\Omega = \Omega_0 \cup \Omega_1 \cup \Omega_2 \cdots \cup \Omega_n \tag{4}$$

其中,$\Omega_i = \{\boldsymbol{x} = (x_1, x_2, \cdots, x_n) \mid \sum_{j=1}^{n} x_j = i\}$,显然,对于任意 $i, j (i \neq j)$,$\Omega_i \cap \Omega_j = \varnothing$。

最大团问题的目标是找到尽可能大的团,如果已经找到一个团数为 l 的极大团,则可以停止搜索子空间 $\Omega_i (1 \leqslant i \leqslant l)$。此外,如果花费了很长的时间在某个子空间中找到 l,那么对于搜索空间 $\Omega_i (i \gg l)$,找到比 l 大的团将会花费更大的代价,对 $\Omega_i (i \gg l)$ 可以不再进行搜索。

这里,将 $n+1$ 个子搜索空间组合成若干阶段,初始化时算法被赋予一个已经确定的团数下界 *bound*,则第一阶段的

收稿日期:2008 - 06 - 11;**修回日期**:2008 - 07 - 25。

作者简介:吴冬晖(1984 -),男,上海人,硕士研究生,主要研究方向:运筹与决策、智能计算; 马良(1964 -),男,上海人,教授,博士生导师,博士,主要研究方向:系统工程、智能计算。

搜索空间为 $\bigcup_{i=bound+1}^{bound+\Delta}\Omega_i$，$\Delta$ 是一个正整数，相当于步长。在某个阶段中若找到一个团数为 s 的极大团，则可进入下一阶段，搜索空间更新为 $\bigcup_{i=s+1}^{s+\Delta}\Omega_i$。算法重复迭代这个过程，随着被找到的团数的增长，单个阶段内的搜索时间也会增加，找到更优解将更加困难。

2.3　启发式局部搜索

任何一个编码 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)\in\{0,1\}^n$ 都表示一个顶点子集 $U\subset V$，由于随机选择的顶点集 U 不一定构成团，通过局部搜索可以对 U 进行修正，并获得团。

本文用以下启发式算法进行局部搜索，找到每个阶段的极大团。

步骤 1

①令 $W:=U,S:=U$；

②若 $W=\varnothing$，转向步骤2；否则从 W 中任意选取一个顶点 k，并将其从 W 中删除；

③ 以 $\alpha\in(0,1)$ 的概率选择将顶点 k 从 S 中删除；否则从 S 和 W 中删除所有在 S 中与顶点 k 不相邻的顶点；

④转向步骤 2。

步骤 2

①令 $W:=V\backslash S$；

②若 $W=\varnothing$，停止并输出 S；否则从 W 中任意选取一个顶点 k，并将其从 W 中删除；

③ 若顶点 k 与 S 中的所有顶点都相邻，将顶点 k 加入到 S 中；

② 转向步骤 2. ②。

经过步骤 1 得到团 S，步骤 2 将 S 扩展为极大团。

2.4　变异操作

设编码 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)\in\{0,1\}^n$ 对应的变异率向量为 $\boldsymbol{p}=(p_1,p_2,\cdots,p_n)\in[0,1]^n$，对于编码 x 的每一位 x_i，变异算子以 β 的概率选择是否进行变异，若变异则以 p_i 的概率选择置为 1，否则置为 0。

在对每一个阶段开始新的搜索时，使用式(1)对变异率向量 $\boldsymbol{p}$ 进行初始化：

$$p_i(t)=\left(\sum_{j=1}^{N}x_i^j\right)/N \tag{5}$$

其中，t 为演化代数，N 为总群大小，$x^j=(x_1^j,x_2^j,\cdots,x_n^j)$，$j=1,2,\cdots,N$ 表示第 t 代的个体，式(1)实际上表示了编码 $\boldsymbol{x}$ 所有位置上的均值，即位置信息。在每个演化代，一些适应性强的个体会被选择作为下一代的父体，假设选择 M 个适应性强的个体作为下一代演化的父本，之后再对 $\boldsymbol{p}$ 进行以下更新：

$$p_i(t)=(1-\lambda)p_i+\lambda\left(\sum_{j=1}^{M}y_i^j\right)/M \tag{6}$$

其中，$y^j=(y_1^j,y_2^j,\cdots,y_n^j)$，$j=1,2,\cdots,M$ 表示 M 个父体，$\lambda\in(0,1]$ 为常数，λ 越大，表示引导个体向适应性强的父体学习，λ 可以被称为学习率，是全局信息。在变异过程中，算法通过记录下优良个体的位置信息，在每一步迭代后以该位置信息更新变异率向量，引导搜索向着更优的方向发展。

2.5　算法描述

1）初始化参数：种群规模 N，搜索空间步长 Δ、α、β、λ；

2）设演化代 $t=0$，随机选择任一 $\boldsymbol{x}\in\boldsymbol{\Omega}$，使用局部搜索找到一个极大团 U，设置搜索下界 $bound=|U|$；

3）从搜索空间 $\bigcup_{i=bound+1}^{bound+\Delta}\Omega_i$ 中随机选取 N 个编码，应用局部搜索分别找到极大团，N 个搜索结果的编码构成了第 t 代的总群，接着以式(1)初始化变异率向量 $\boldsymbol{p}$；

4）从第 t 代中选择 $N/2$ 个较优的个体构成父体，并以式(2)更新向量 $\boldsymbol{p}$；

5）对第 t 代父体中的一个最优个体进行 $N/2$ 次变异操作，并进行局部搜索，得到 $N/2$ 个新解，这 $N/2$ 个新解同第 t 代父体一起构成了第 $t+1$ 代的总群。这里，若满足终止条件，则停止搜索并输出当前最优解；

6）令 $t=t+1$，设团 S 为第 t 代总群中的最优解，如果 $|S|>bound$，令 $bound=|S|$，并转 3)；

7）若第 t 代的种群中所有的个体都相同，则转3)，否则转4)。

3　算法实例测试

将本文算法在 Matlab 6.5 上进行了实现，并在 Intel 2.1 GHz环境下测试，测试实例为 DIMACS 基准图 C1000.9（顶点数 1 000，边数 450 079），该实实例可以从 http://dimacs.rutgers.edu/Challenges/上获得。

实例测试中的固定参数设置如下：种群规模 $N=10$，步长 $\Delta=3$，$\alpha=0.001$，此外，变异率 β、学习率 λ 的参数值将直接影响算法的性能。为了减少 β、λ 对算法性能的影响，本文分别以 $\beta=0,0.1,0.2,\cdots,1$，$\lambda=0,0.1,0.2,\cdots,1$ 对C1000.9 进行了测试，结果如图 1。

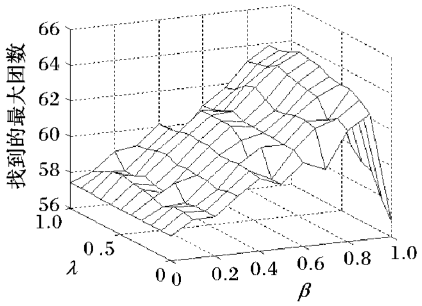

图 1　C1000.9 算例在不同 β、λ 组合下找到的最大团数

图 1 解释了两个参数 β、λ 对算法性能的作用，可以看到，在 $\beta<0.9$ 之前，随着 β 值的增加，算法的性能也逐步增强，这说明了变异率向量 $\boldsymbol{p}$ 中的位置信息对算法起着积极的作用；而当 $\beta=1$ 时，表示任何情况下都使用位置信息进行变异操作，算法陷入了局部最优，结果并不理想。当学习率 $\lambda=0$，即不使用全局的学习信息引导变异，算法的表现较差，而当 $\lambda\in[0.5,0.7]$ 时，算法性能又有所提高，因此这个区域是一个较好的参数配置。通过对 C1000.9 的实例测试，证明本文提出的最大团问题的求解算法是一个有效算法，且满足求解中等规模图的需求。

参考文献：

[1]　王丽爱，周旭东，陈崚. 最大团问题研究进展及算法测试标准[J]. 计算机应用研究，2007，24(7)：69－71.

[2]　刘怀义，杨小帆，孙丽萍，等. 用带非线性自反馈的神经网络求解最大团问题[J]. 重庆大学学报：自然科学版，2007，30(9)：60－63.

[3]　谢政. 网络算法与复杂性理论[M]. 长沙：国防科技大学出版社，2003.

[4]　PARDALOS P M, RAPPE J, RESENDE M G C. An exact parallel algorithm for the maximum clique problem[C]// High Performance Algorithms and Software in Nonlinear Optimization. [S. l.]: Kluwer Acedemic Publishers, 1998: 279－300.

[5]　周旭东，王丽爱，陈崚. 启发式算法求解最大团问题研究[J]. 计算机工程与设计，2007，28(18)：4329－4332.