

求解 k 最短路径问题的混合遗传算法

赵礼峰, 于汶雨

(南京邮电大学 理学院, 江苏 南京 210023)

摘要:遗传算法求解问题的关键在于对问题的解进行编码,同时需要构造出适应度函数。结合 k 最短路径实际问题,重新定义了一种染色体编码方式,并且新构造了符合该问题的适应度函数。标准遗传算法采用固定的交叉率和变异率,在应用过程中存在收敛过慢、早熟及稳定性差的缺点。因此,提出了一种改进的自适应遗传算法,对交叉率和变异率采用自适应方式,构造了确定交叉率和变异率的公式,加快了算法收敛速度。同时结合模拟退火的 Metropolis 准则对子代个体的接收做出选择,克服了算法容易早熟的问题。仿真结果表明,改进后的混合遗传算法可以求解 k 最短路径问题,并且在寻优精确度、时间效率、稳定性上均优于标准遗传算法。

关键词:混合遗传算法;染色体编码;Metropolis 准则; k 最短路径

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2016)10-0032-04

doi:10.3969/j.issn.1673-629X.2016.10.007

A Hybrid Genetic Algorithm for Solving k Shortest Path Problem

ZHAO Li-feng, YU Wen-yu

(College of Sciences, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract: The key for the genetic algorithm is coding and the need to construct the fitness function. Combined with actual k shortest path problem, a new chromosome coding method is redefined and a new fitness function is constructed. The standard genetic algorithm adopts fixed crossover rate and mutation rate, and exists the defects of low convergence, prematurity and poor stability. Therefore, an improved adaptive genetic algorithm is proposed, using the adaptive way for the crossover rate and mutation rate, the formula for determining them is construct to accelerate the convergence speed. At the same time, combined with the Metropolis principle of simulated annealing to choose the receiver of the offspring, the algorithm overcomes the problem of premature. The simulation shows that the improved hybrid genetic algorithm can solve the k shortest path problem, and it is superior to the standard genetic algorithm in optimization accuracy, time efficiency and stability.

Key words: hybrid genetic algorithm; chromosome code; Metropolis principle; k shortest path

0 引言

k 最短路径问题^[1]是图论中研究的一个重要课题,同时在现代计算机网络以及交通系统中扮演着重要的角色。该问题最早由 Hoffman 和 Pavley 在 20 世纪 50 年代提出^[2],多年来也一直受到学术界的广泛关注,并且先后提出了多种求解 k 最短路径问题的算法。因此,如何快速求解 k 最短路径问题成为了研究的重点。文献[3]主要的研究成果是基于 Dijkstra 算法的前 N 条最短路径算法;文献[4-6]利用“偏离路径”的性质给出计算 k 条最短路径的递推公式,但在求解第 k 条最短路径时采用了不同的方法。

遗传算法 (Genetic Algorithm)^[7]是由美国 J. Holland 教授在 1975 年首先提出的、模拟达尔文生物进化论的自然选择和遗传学机理的生物进化过程的计算模型,是一种通过模拟自然进化过程演化而来的随机化搜索方法。近年来在世界范围内形成进化计算热潮,计算智能已成为人工智能研究的一个重要方向,加上后来兴起的人工生命研究,使得遗传算法受到了极大的关注。

遗传算法虽然具有良好的全局搜索能力,可以快速地搜索出解空间中的全体解,但算法存在收敛速度慢或易出现“早熟”现象等缺陷。文中针对遗传算法

收稿日期:2015-10-25

修回日期:2016-03-03

网络出版时间:2016-09-18

基金项目:国家自然科学基金资助项目(61070234,61071167)

作者简介:赵礼峰(1959-),男,教授,硕士研究生导师,研究方向为图论及其在通信中的应用;于汶雨(1989-),女,硕士研究生,研究方向为图论及其在通信中的应用。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160918.1707.020.html>

不足之处进行改进。为克服遗传算法中 p_c 和 p_m 不好确定的缺点,对 p_c 和 p_m 采用自适应过程。结合模拟退火算法的 Metropolis 准则,提高局部搜索能力,克服遗传算法早熟问题。同时为使遗传算法使用于 k 最短路径问题,构造一种新的染色体编码方式和染色体的交叉操作。仿真结果表明,该算法可以提高运行效率,具有较高的精确度并且结构简明。

1 预备知识

设有赋权图^[8] $G = (V(G), E(G))$, 其中, V 为 n 个节点的集合, E 为 m 条边的集合, 简单图 G 中每条边的权值为 $w(e)$ 。图 G 中从点 s 到点 t 的路径 p 表示为 $P_c(s, t) = v_s e_1 v_2 e_2 \cdots e_l v_t$, 同时路径权值记为 $d_G(s, t) = \sum_{i=1}^l w(e_i)$ 。

那么对于给定的点 s 和点 t 之间的 k 最短路径问题就是在赋权图 G 中求出点 s 到点 t 的 k 条路 P_1, P_2, \dots, P_k , 同时需要满足条件 $W(P_1) \leq W(P_2) \leq \dots \leq W(P_k)$ 。

2 遗传算法和模拟退火算法

2.1 遗传算法基本步骤

遗传算法寻优过程其实是一个迭代的过程。把搜索的解空间映射为遗传的空间, 将解映射成为染色体, 向量中的每个元素称为基因, 将所有的染色体组成种群, 并按目标函数对每个染色体的优劣进行评价, 并根据结果给出一个适应度值 (fitness)。在这种情况下, 每一代中各个体的特征可通过染色体遗传到下一代中。在下一代中, 一个群体中合体相互之间可以复制 (reproduction) 和交叉 (crossover), 并会以一定的概率发生变异 (mutation)。交叉倾向于选择群体中最为优秀的个体, 这些交叉个体的最优特性相结合后产生的后代比父代具有更优良的特性, 从而产生较好的解。遗传算法基本流程^[9]如图 1 所示。

2.2 模拟退火基本原理与 Metropolis 准则

模拟退火算法 (Simulated Annealing, SA) 是模拟热力学中经典粒子系统的降温过程来求解规划问题的极值。当孤立粒子系统的温度以足够慢的速度下降时, 系统近似处于热力学平衡状态, 最后系统将达到本身的最低能量状态, 即基态, 这相当于能量函数的全局极小点。

模拟退火算法的基本原理如下:

(1) 给定初始温度 T_0 及初始点, 计算该点的函数值 $f(x)$ 。

(2) 随机产生扰动 Δx , 得到新点 $x' = x + \Delta x$, 计算

新点函数值 $f(x')$ 及函数值差 $\Delta f = f(x') - f(x)$ 。

(3) 若 $\Delta f \leq 0$, 则接受新点, 作为下一次模拟的初始点。

(4) 若 $\Delta f > 0$, 则计算新点接受概率: $p(\Delta f) = \exp\left(-\frac{\Delta f}{K \cdot T}\right)$ 。产生 $[0, 1]$ 区间上均匀分布的伪随机数 r , $r \in [0, 1]$, 如果 $p(\Delta f) \geq r$, 则接受新点作为下次模拟的初始点; 否则放弃新点, 仍选取原来的点作为下次模拟的初始点。

以上步骤称为 Metropolis 准则^[10]。按照一定的退火方案逐渐降低温度, 重复 Metropolis 过程, 就构成了模拟退火优化算法。

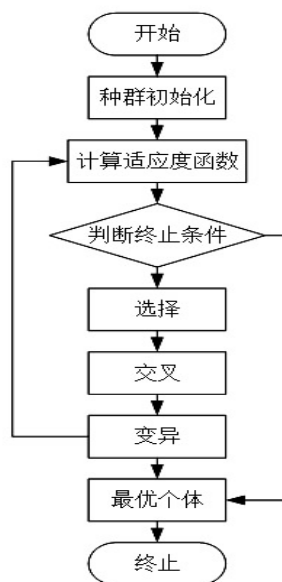


图 1 遗传算法流程图

3 算法改进

3.1 染色体编码

结合 k 最短路径问题实际情况, 文中定义了一种新的染色体编码方式。

假设在无向赋权图 $G = (V, E)$ 中, 初始节点 V_s 到目的节点 V_t 的路径为: $1 \rightarrow 5 \rightarrow 9 \rightarrow 13 \rightarrow 17 \rightarrow 20$ 。其中, 每个数字代表的是节点编号, 在遗传算法中可以作为每个染色体的基因, 不同的路径包含有不同的基因点, 所以设计染色体的编码方式为根据节点编号编码, 上述路径的染色体编码为: $[1-5-9-13-17-20]$ 。

3.2 适应度函数和选择概率

适应度函数即为计算群体 $P(t)$ 中每条染色体的适应度。遗传算法与 k 最短路径问题相结合构造的适应度函数为:

$$\text{Fitness} = \frac{1}{\sum_{e \in P} w(e)} \quad (1)$$

其中, P 为图 $G=(V,E)$ 中的任意一条路径,即为遗传算法中的每条染色体; e 为路径 P 上的每条边; $w(e)$ 为每条边的权值。

选择概率与其适应度呈比例,假设群体大小为 N ,第 i 个染色体的适应度为 f_i ,第 i 个染色体被选中的概率为 P_i ,那么

$$P_i = \frac{f_i}{\sum_{i=1}^N f_i} \quad (2)$$

显然个体适应度越大被选择的概率越高。

3.3 交叉操作与变异操作

交叉操作为两个父代个体的部分结构替换重组从而生成新个体的操作。文中采用两点交叉的方式,根据 k 最短路径问题定义一种新的交叉方式,如下所示:

父代 1:

[1 - 3 - 7 - 10 - 5 - 9 - 11 - 14 - 18 - 20]

父代 2:

[1 - 2 - 6 - 7 - 13 - 10 - 8 - 9 - 16 - 19 - 20]

选取两个基因点: 7 与 9, 交换两个基因之间的部分即完成交叉,交叉后为:

子代 1:

[1 - 3 - 7 - 13 - 10 - 8 - 9 - 11 - 14 - 18 - 20]

子代 2:

[1 - 2 - 6 - 7 - 10 - 5 - 9 - 16 - 19 - 20]

由此可见,两个父代个体在两个交叉点之间的码串互相交换,分别生成两个新的子代染色体。此外,若在交叉操作后产生环路,例如子代染色体为 [1 - 4 - 2 - 5 - 4 - 9], 含有相同的基因点,则剔除公共部分,编码变为 [1 - 4 - 9]。

变异操作的基本内容是对群体中的染色体个体上的某些基因点的基因值进行变更。其目的是可以使遗传算法具备局部随机搜索能力和保持群体的多样性,避免出现未成熟就收敛的情况。在最短路径问题中,选择要变异的路径,假设为 Y :

$Y: [V_s a_1 a_2 a_3 a_4 a_5 a_6 V_t]$

随机选取 Y 上的基因块,例如选取 $(a_2 a_3 a_4)$, 此时随机生成以 a_2 为起点 a_4 为终点的路径 $(a_2 X a_4)$ 为新基因块取代选取出的基因块,其中 X 代表所有经过的基因点,变异后的染色体为:

$Y': [V_s a_1 a_2 X a_4 a_5 a_6 V_t]$

若变异后产生环路,按照交叉操作中所述进行处理。

3.4 交叉率(p_c)与变异率(p_m)的确定

遗传算法在进化过程中,进化程度与交叉率和变

异率^[11]的取值有很大关系, p_c 取值越大,新个体产生的速度越快,但当 p_c 过大时又会使高适应度的个体结构快速被破坏,当 p_c 取值过小时会使搜索过程缓慢,甚至停滞不前。当变异概率 p_m 取值过小时,不易产生新的结构;当 p_m 取值过大时,遗传算法就会变成纯粹的随机搜索算法。

所以文中采取自适应的遗传算法。根据交叉率和变异率对遗传算法群体进化过程中的影响,设计如下公式进行自适应调整:

$$p_c = \begin{cases} p_{c_1} - p_{c_2} \frac{f_c - f_{avg}}{f_{max} - f_{avg}} & f_c \geq f_{avg} \\ p_{c_1} & f_c < f_{avg} \end{cases} \quad p_{c_1} = 0.9, p_{c_2} = 0.2 \quad (3)$$

$$p_m = \begin{cases} p_{m_1} - p_{m_2} \frac{f_c - f_{avg}}{f_{max} - f_{avg}} & f_c \geq f_{avg} \\ p_{m_1} & f_c < f_{avg} \end{cases} \quad p_{m_1} = 0.01, p_{m_2} = 0.05 \quad (4)$$

其中, p_c 为两个交叉父代个体中适应度值较大的个体; f_{avg} 为每一代群体的平均适应度值; f_{max} 为群体中适应度值最大的个体。

当个体的适应度值小于平均值时,采用较高的 p_c 和 p_m ,使个体进化成为优秀个体的概率增大;当个体的适应度值大于平均适应度值时,采用较低的 p_c 和 p_m ,这样可以确保优秀个体的结构不容易被改变。

3.5 对最优个体进行模拟退火操作

模拟退火算法^[12]的优点之一就是能以一定的概率接收较差解。也就是说算法即使陷入局部最优,理论上经过足够长的时间也可跳出来,从而收敛到全局最优解。新解是否被接受,判断的依据为 Metropolis 准则:

$$p = \begin{cases} 1 & f(i) \leq f(i') \\ \exp\left[\frac{f(i) - f(i')}{T_k}\right] & f(i) > f(i') \end{cases} \quad (5)$$

文中采用快速降温策略,其中 $T_k = \frac{T_0}{1+k}$, T_k 代表退火温度, T_0 代表初始温度, k 代表迭代次数, f 代表个体适应度函数。

经过交叉操作后,若子代个体适应度变优,完全接收新解为当前解;若子代个体恶化,则以概率 p 接受恶化解为新的当前解。同时为了使每一代中优良个体结构不被破坏,采用精英保留策略^[13],如果下一代群体中的最优个体适应度值小于当前群体最优个体适应度值,那么用当前群体中最优个体或者适应度值大于下一代最优个体适应度值的多个个体随机替代下一代群体中适应度差的相应数量的个体,确保了当前的最优个体不会因为交叉、变异等操作而被破坏。

3.6 算法流程

- Step1:群体初始化,随机选取 n 条路径构成群体;
- Step2:对群体中的每一个个体,根据式(1)和式(2)分别计算适应度值和选择概率;
- Step3:按照 Step2 中得到的选择概率选择需要交叉操作的个体;
- Step4:根据式(3)确定交叉率,根据 3.3 中所述的交叉方式进行交叉操作;
- Step5:根据式(4)确定变异率,根据 3.3 中所述的变异方式进行变异操作;
- Step6:按照模拟退火 Metropolis 准则对交叉后的个体是否接收进行操作;
- Step7:根据精英保留策略对子代群体进行操作;
- Step8:判断是否达到终止条件,若达到终止条件则输出前 k 个优秀解,否则转 Step2。

4 仿真实验

仿真实验环境:处理器为 Intel COREi5,操作系统为 Windows 7,应用 MATLAB^[14] 编写。以图 2 为算例,其中共有 30 个网络节点,对算法的可行性及精确度进行实验。

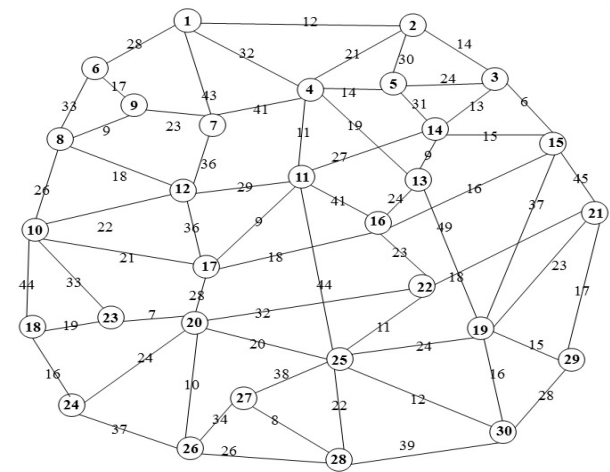


图 2 网络拓扑图

当 $k = 10$ 时,标准遗传算法(GA)与改进后遗传算法(GAM)所求 v_1 到 v_{30} 路径权值对比如图 3 所示。

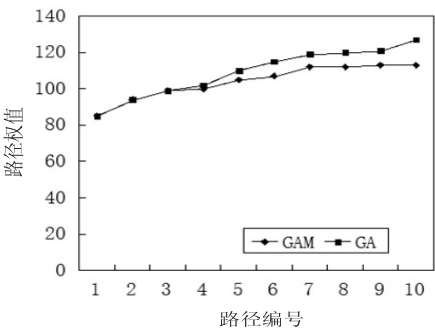


图 3 求解结果精度对比

其中,初始种群个体数量 $n = 100$,迭代次数定为 5 000, p_c 与 p_m 由自适应公式得出,初始温度 $T_0 = 200$ 。从图中可以看出,改进后的算法具有更高的精确度。

图 4 为两种算法在不同规模网络中求解时间的比较。起始求解网络规模为 20 个节点,150 条边,每次求得结果后,网络拓扑图的节点个数和边的条数以 5, 50 的数量递增,直到增加到 50 个节点、450 条边为止。

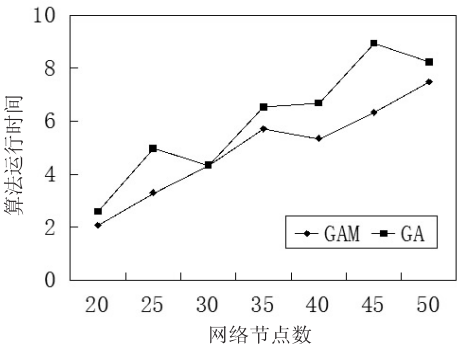


图 4 两种算法求解时间对比

从图中可以看出, GAM 平均时间效率高 于 GA ,同时曲线更加平缓,说明稳定性也高于 GA 。

5 结束语

通过把遗传算法与模拟退火算法相结合,文中提出了一种求解 k 最短路径问题的新算法。通过对交叉率与变异率采用自适应的方式,构造了新的染色体编码方式及适应度函数,实现了对 k 最短路问题的求解。仿真结果表明,该算法能够较快较准确地找到多条最短路径,具有较高稳定性的同时克服了遗传算法易早熟的缺点。对求解一般复杂网络中 k 最短路径问题,具有广泛的应用价值。

参考文献:

[1] Palmgren M, Yuan D. A short summary on K shortest path: algorithms and applications [EB/OL]. 1999. <http://www.esc.auckland.ac.nz/Mason/Courses/LinkopingColCen99/kth.pdf>.
[2] Hoffman W, Pavley R. A method of solution of the Nth best path problem[J]. Journal of the ACM, 1959, 6(4): 506-514.
[3] 柴登峰,张登荣.前 N 条最短路径问题的算法及应用[J]. 浙江大学学报:工学版, 2002, 36(5): 531-534.
[4] 袁红涛,朱美正. K 优路径的一种求解算法与实现[J]. 计算机工程与应用, 2004, 40(6): 51-53.
[5] Kang T, Zhang X, Wang Z, et al. Algorithm for shortest path of multi-objectives based on k short path algorithm[J]. Journal of Changzhou Institute of Technology, 2011, 24(3): 25-28.
[6] Wang Z, Han W, Li Y. Shortest path problem with multiple shortest paths[J]. Journal of Harbin Institute of Technology,

显更小,能有效降低搜索流量,同时也能保证一定的资源搜索量。

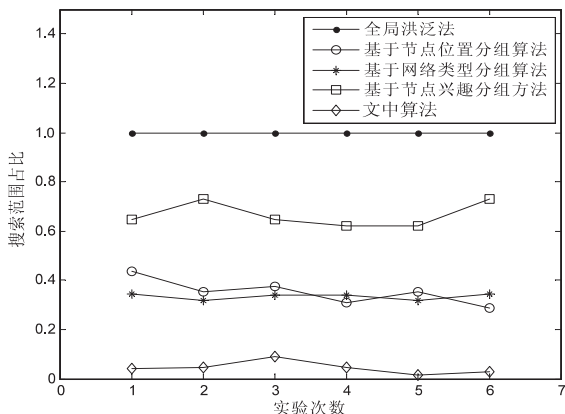


图 6 搜索范围占比

6 结束语

文中提出一种混合流媒体分发系统中的多指标用户群分组算法,并通过仿真实验验证了该算法的有效性,但对于其在真实网络环境中的应用价值尚有待验证。下一步的研究将集中于此,将该算法部署到流媒体分发系统中加以论证。

参考文献:

- [1] Ratnasamy S, Handley M, Karp R, et al. Topologically-aware overlay construction and server selection [C]//Proc of INFOCOM 2002. [s. l.]: IEEE, 2002: 1190–1199.
- [2] 周振朝, 费耀平, 李敏. 基于网络坐标的无结构 P2P 节点聚类算法[J]. 计算机工程, 2010, 36(11): 98–100.
- [3] 邢长友, 陈鸣. 网络距离预测技术[J]. 软件学报, 2009, 20(9): 2470–2482.
- [4] 王意洁, 李小勇. 网络距离预测技术研究[J]. 软件学报, 2009,

20(6): 1574–1590.

- [5] 吴思, 欧阳松. 基于兴趣相关度的 P2P 网络搜索优化算法[J]. 计算机工程, 2008, 34(11): 102–104.
- [6] 赵捧未, 马琳, 秦春秀. P2P 用户兴趣社区形成研究[J]. 现代图书情报技术, 2013(10): 53–58.
- [7] 梁卫芳, 黄建华. 基于网络拓扑和节点兴趣的 P2P 搜索机制[J]. 计算机工程与设计, 2008, 29(6): 1316–1318.
- [8] 郑力明, 李晓冬, 李小勇, 等. P2P 覆盖网中的聚类研究综述[J]. 计算机应用研究, 2010, 27(3): 806–810.
- [9] 杨博, 刘大有, LIU Jiming, 等. 复杂网络聚类方法[J]. 软件学报, 2009, 20(1): 54–66.
- [10] 孙吉贵, 刘杰, 赵连宇. 聚类算法研究[J]. 软件学报, 2008, 19(1): 48–61.
- [11] Adhikari V K, Jain S, Chen Y, et al. Vivisecting YouTube: an active measurement study [C]//Proc of INFOCOM. Orlando, FL: IEEE, 2012: 2521–2525.
- [12] Adhikari V K, Guo Y, Hao F, et al. Unreeling netflix: understanding and improving multi-CDN movie delivery [C]//Proc of INFOCOM. Orlando, FL: IEEE, 2012: 1620–1628.
- [13] Adhikari V K, Guo Y, Hao F, et al. A tale of three CDNs: an active measurement study of Hulu and its CDNs [C]//Proc of IEEE conference on computer communications workshops. Orlando, FL: IEEE, 2012: 7–12.
- [14] Huang Yan, Fu T Z J, Chiu Dah-Ming, et al. Challenges, design and analysis of a large-scale p2p-vod system [C]//Proc of SIGCOMM. [s. l.]: ACM, 2008: 375–388.
- [15] Lei J, Shi L, Fu X. An experimental analysis of Joost peer-to-peer VoD service [J]. Peer-to-Peer Netw. Appl., 2010, 3(4): 351–362.
- [16] Zhang G, Liu W, Hei X, et al. Unreeling Xunlei Kankan: understanding hybrid CDN-P2P video-on-demand streaming [J]. IEEE Transactions on Multimedia, 2015, 17(2): 229–242.

(上接第 35 页)

- 2010, 42(9): 1428–1431.
- [7] 周明, 孙树栋. 遗传算法原理及应用[M]. 北京: 国防工业出版社, 2001.
- [8] 谢政, 李建平. 网络算法与复杂性理论[M]. 长沙: 国防科技大学出版社, 2003.
- [9] Alsultanny Y A, Aqel M M. Pattern recognition using multilayer neural-genetic algorithm [J]. Neurocomputing, 2003, 51: 237–247.
- [10] 康立山. 非数值并行算法(第一册): 模拟退火算法[M]. 北京: 科学出版社, 2000: 22–38.
- [11] 赵礼峰, 王小龙. 图的 Steiner 最小树问题的混合遗传算法

[J]. 计算机技术与发展, 2014, 24(10): 110–114.

- [12] 鲁建业, 李琦, 董蕴华, 等. 采用混合遗传-模拟退火算法对 DOE 的直接设计[J]. 光电子·激光, 2001, 12(4): 365–367.
- [13] Gao Erbao, Lai M. An improved genetic algorithm for the vehicle routing problem with simultaneous delivery and pickup [C]//Proc of the 6th Wuhan international conference on e-business – innovation management track. Wuhan: [s. n.], 2007: 2100–2106.
- [14] 王海英. MATLAB 遗传算法工具箱及应用[M]. 北京: 北京航空航天大学出版社, 2010.