

组播服务节点选择问题的求解算法

程德强¹, 钱建生¹, 杨红丽²

(1. 中国矿业大学信息与电气工程学院 江苏 徐州 221008; 2. 电子科技大学通信与信息工程学院 成都 610054)

【摘要】为了解覆盖网络中最优组播服务节点MSN组合问题,提出一种基于混合实数编码的遗传算法(HRCGA)。采用实数编码方式对MSN组合进行编码,利用K-medoids聚类距离算子进行节点相似性度量,并根据MSN的实际特点,构造含有惩罚因子的适应度函数,限制不良个体参与进化。同时,针对标准遗传算法SGA局部搜索能力弱的特点,HRGCA引入个体进化控制策略。理论分析和仿真结果表明,该算法有效克服了传统K-medoids算法易陷入局部极小值、对初始中心选值敏感的问题。通过与传统SGA的仿真结果对比, HRCGA进一步提高了MSN组合全局寻优能力

关键词 遗传算法; 混合算法; 覆盖网络; 实数编码

中图分类号 TN919.8 **文献标识码** A

A Solution to Select Multicast Service Nodes of Overlay Network

CHENG De-qiang¹, QIAN Jian-sheng¹, and YANG Hong-li²

(1. School of Information and Electrical Engineering, China University of Mining and Technology Xuzhou Jiangsu 221008;

2. School of Communication and Information Engineering, University of Electronic Science and Technology of China Chengdu 610054)

Abstract In order to select optimal multicast service node (MSN) of layered overlay network, a hybrid real-coded genetic algorithm (HRCGA) is proposed. The fitness function with penalty operator is created to prevent bad chromosomes participating evolution. Also, an evolving control strategy is proposed to enhance the local search ability. Results of computer simulation show that HRCGA not only avoids converging to local minimum value, but also is robust to initialization. And the global search ability is enhanced compared with SGA.

Key words genetic algorithm; hybrid algorithm; overlay network; real-coded

对于实时流媒体数据的传输,覆盖组播网络(overlay multicast network, OMN)^[1]克服了IP网络层组播需要网络中的路由器支持、可扩展性、网络管理等屏障^[2]。

与文献[4-5]等覆盖网络协议相比,文献[3]采用特定源树(source-specified tree)的方式管理网络组播节点,在进行大容量数据转发时,其链路压力最均衡^[6]。本文以NICE网络结构为基础,进行网络节点的管理,构建基于分层的特定源组播树,进行视频数据包的分发。所有节点属于最底层 L_0 ,节点根据相互之间的相似性(距离)形成多个簇,并在簇中选取中心节点(簇头)形成层 L_1 ,然后根据 L_1 各节点的相似性形成多个簇,相应的簇头构成 L_2 ,重复以上步骤,最终形成分层组播基础结构(application level multicast infrastructure, ALMI)。由于各个簇头与簇中成员相似性最高,用来担负主干数据流的转发,构成了基于特定源分发组播树。其中,各个簇头称

之为组播服务节点(MSN),各个簇称为组播岛(multicast island, MI)。

覆盖网络中,组播转发树的建立要尽可能的减少网络延时,同时保证网络负载的均衡,其中MSN的选取非常重要。因此,本文主要解决的就是结合视频图像传输的特点,提出MSN全局最优选取的方法。为简化起见,只考虑 L_0 层组播服务节点MSN组合的选取问题,其方法同样适用其他层MSN组合的选取。

考虑到网络节点为离散的实际点,并依据K-medoids算法具有较强的局部搜索能力和对孤立点具有很好的健壮性的特点,结合遗传算法的高效率全局优化能力,采用混合实数编码机制^[7-8],本文提出基于K-medoids和遗传算法结合的混合算法HRCGA,以此进行MSN组合的选取,达到构建基于固定节点配置的分层组播覆盖网络的目的。同时,针对标准遗传算法SGA的早熟收敛现象^[9],根据覆

盖网络MSN组合的特点,构造含有惩罚因子的适应度函数;HRGCA引入个体进化控制策略,对SGA进行优化改进。一方面限制不良个体生成,缩小搜索空间,防止较优个体在进化中的扩散;另一方面,加快收敛速度,提高算法在解空间的寻优能力。最后通过仿真实验,验证了算法的有效性。

1 MSN组合选取问题的数学模型

覆盖组播网络中MSN组合的最优选取目标是组播岛中各节点到所属MSN的成本和最小。组播岛中各节点到所属MSN的成本由各节点到MSN的总链路延迟和MSN的转发能力决定,其中MSN的转发能力代表可向下游节点提供视频连接通路的能力,其受MSN节点的处理速度和所在网络状况影响。

在MSN组合的全局最优选取中,根据节点与MSN的距离相似度,利用最近邻法则,使节点距离所属组播岛MSN的距离最近。因此,MSN组合优化选择的目标函数为:

$$\min J = \sum_{h=1}^c \sum_{j=1}^n d(s_j, v_h) \quad (1)$$

式中 J 为组播岛中各节点到所属MSN的成本(目标函数),用类内平方误差和WGSS表示; c 为覆盖组播网络中被选中为MSN的节点个数; n 为网络中所有节点个数; s_j 、 v_h 分别是网络节点和MSN的特征向量。 $d(s_j, v_h)$ 为 s_j 到 v_h 的K-medoids聚类距离算子,代表 s_j 到 v_h 距离成本。根据覆盖网络特点,组播岛中各节点到所属MSN的成本由组播岛各节点到对应MSN的总链路延迟和MSN的转发能力决定,因此, $d(s_j, v_h)$ 可定义为:

$$d(s_j, v_h) = u_{hj} [w_1^2 a_{hj}^2 + w_2^2 r_h^2] \quad (2)$$

式中 w_1 和 w_2 为节点特征向量中各分量在距离成本中的权重; a_{hj} 为 j 节点到序号为 h 的MSN之间的链路延时; r_h 为序号为 h 的MSN节点转发能力。

MSN组合优化转变为求解式(1)目标函数达到最小时MSN组合的选择问题,并且其约束条件为:

$$u_{hj} = \begin{cases} 1 & d_{hj} = \min\{d_{kj}\} \\ 0 & d_{hj} \neq \min\{d_{kj}\} \end{cases} \quad (3)$$

$$w_1 + w_2 = 1 \quad (4)$$

其中,式(3)为硬聚类划分约束;式(4)为节点特征属性权重约束。

2 混合实数编码遗传算法求解

2.1 MSN组合选择问题的矩阵实数编码

利用网络节点间链路RTT作为节点间的距离度

量,则令 n 阶差异矩阵 A 表示 n 个节点间的链路RTT;用 n 维行向量 $R = (r_1, r_2, \dots, r_i, \dots, r_n)$ 表示节点的视频数据转发处理能力,取 $1 \leq r_i \leq 10$, r_i 越小表示MSN节点的视频数据转发处理能力越强。对网络节点的特征属性,采用矩阵形式进行实数编码,其特征属性矩阵 P 具体形式为:

$$P = [p_1, p_2, \dots, p_k, \dots, p_n] = \begin{bmatrix} A \\ R \end{bmatrix} = \begin{bmatrix} 0 & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & 0 & \dots & \vdots \\ \vdots & \dots & \ddots & a_{n-1,n} \\ a_{n,1} & \dots & \dots & 0 \\ r_{n+1,1} & r_{n+1,2} & \dots & r_{n+1,n} \end{bmatrix} \quad (5)$$

式中 $p_k = (a_{1,k}, a_{2,k}, \dots, a_{j,k}, \dots, a_{n,k}, r_{n+1,k})^T$ 为 k 节点的特征矢量; $a_{j,k}$ ($1 \leq j \leq n$) 为 k 节点与其他节点的RTT距离度量; $r_{n+1,k}$ 为 k 节点的视频数据处理转发能力。并且,根据差异矩阵的特点,则:

$$\begin{cases} a_{j,k} = a_{k,j} & k \neq j \\ a_{j,k} = 0 & k = j \end{cases} \quad (6)$$

2.2 遗传种群初始化

遗传种群初始化过程如下:

(1) 按照MSN组合中的MSN个数 c ,从实数编码矩阵 P ,随机选择 c 个列向量,构成一个初始个体,代表一组聚类中心。

(2) 重复操作(1),生成maxPop个初始个体。

2.3 适应度函数选取

组播岛中的对象应偏离MSNs最小,根据式(2),以WGSS为聚类目标函数,目标函数越小,则聚类质量越好,相应的个体适应度越高,则第 k 个个体的目标函数 $J(k)$ 为:

$$J(k) = \sum_{h=1}^c \sum_{j=1}^n d(s_j, v_h) = \sum_{h=1}^c \sum_{j=1}^n u_{hj} [w_1^2 a_{hj}^2 + w_2^2 r_h^2] \quad (7)$$

根据MSN的实际特点,每个MSN代表一个组播岛中心,各中心应该互不相同,因此,个体中基因不可能存在重复。为了防止个体进化过程中此类不良个体参与进化,构造含有惩罚项的适应度函数为:

$$f(k) = \mu / (1 + J(k)) \quad (8)$$

式中 μ 为惩罚乘子。如果个体中存在重复基因,则令 $\mu = 0$; 否则,令 $\mu = 1$ 。

2.4 遗传操作

2.4.1 选择算子

选择算子采用赌轮法获取,则第 i 个个体的选择概率为:

$$p(V_i) = f(i) / \sum_{k=1}^n f(k) \quad i=1,2,\dots,n \quad (9)$$

2.4.2 交叉算子

为了维护群体多样性, 采用近邻配对原则对个体实现交叉, 此种交叉方法可避免较优模式过快的扩散, 而且还符合基因算法细粒度并行模型的要求, 易于获得较大的并行度^[10]。由于覆盖网络, 个体所包含的基因长度有限, 因此本文采用整体算术交叉相互交换两个配对个体的基因。

若个体 $p(V_i) < p_c$ (p_c 为交叉概率), 按照随机配对原则选出的父代个体为 parent(1) 和 parent(2), 其子代个体分别为 child(1) 和 child(2), 并用 parent(i)·(j) 表示第 i 个父代个体的第 j 个基因, child(i)·(j) 表示第 i 个子代个体的第 j 个基因, 利用整体算术交叉, 可得:

$$\text{child}(1) \cdot (j) = \text{round}(\alpha * \text{parent}(1)(j) + (1 - \alpha)\text{parent}(2)(j)) \quad (10)$$

$$\text{child}(2) \cdot (j) = \text{round}(\alpha * \text{parent}(2)(j) + (1 - \alpha)\text{parent}(1)(j)) \quad (11)$$

式中 $\alpha(0 < \alpha < 1)$ 为交叉因子。

2.4.3 变异算子

通过变异运算使得遗传算法保持种群多样性的同时, 具有局部的随机搜索能力。

若个体 parent(i) 满足 $p(V_i) > p_m$ (p_m 为变异概率)采用单点变异, 并把参与变异的分量作随机扰动, 即:

$$\text{child}(i) \cdot (j) = \text{parent}(i)(j) + \delta \quad (12)$$

式中 $\delta = \text{round}(\lambda * \beta)$, β 为 $(-6, 6)$ 区间的均匀分布随机变量, $\lambda = K / t$, K 为正整数, t 为迭代次数。由式(12)可知, 扰动在进化初期变化范围较大, 随着进化代数的增加扰动变化范围逐渐减小, 以此加强进化后期的局部搜索能力。

2.5 个体进化控制策略

HRCGA在遗传操作中, 采用实数矩阵编码, 通过整体算术交叉的方式对个体进行高概率交叉, 以此加快新个体产生的速度, 保证群体多样性, 但对进化过程中的较优模式, 破坏概率也较高。因此, 需增强算法的局部搜索能力, 本文选择高变异概率进行个体单点非一致性变异。传统的全种群变异策略认为变异概率 p_m 不能太大^[10], 否则高频度的变异将使遗传算法趋于纯粹的随机搜索。而本文在变异操作中, 采用个体进化控制策略, 如果变异后的个体产生退化现象, 则取消此次变异, 保留父个体; 否则, 进行变异, 因此, 即使变异概率很大, 不会

破坏优良个体, 也可以保证解的收敛性, 同时在变异中引入随机扰动, 并使扰动范围随进化代数减小, 进一步提高进化后期的局部搜索精度, 仿真实验也验证了其有效性。

3 仿真实验

随机生成10组网络节点样本集, 每组150个节点。对于每组样本集, 随机生成初始种群, 令其种群个体数目为30, 每个个体中有5个基因, 即5个组播服务节点。

3.1 全局寻优能力分析

分别利用HRCGA、SGA和K-medoids选择算法进行组播服务节点MSN组合的选取。根据聚类的特点, 达到稳定最优解时, 目标函数越小, 则MSNs选择性能越好。3种选择算法针对10组网络节点样本集, 选择出最优解时的最小目标函数, 如图1所示。其中, J_{\min} 为最优解对应的目标函数; S 为网络节点样本集。

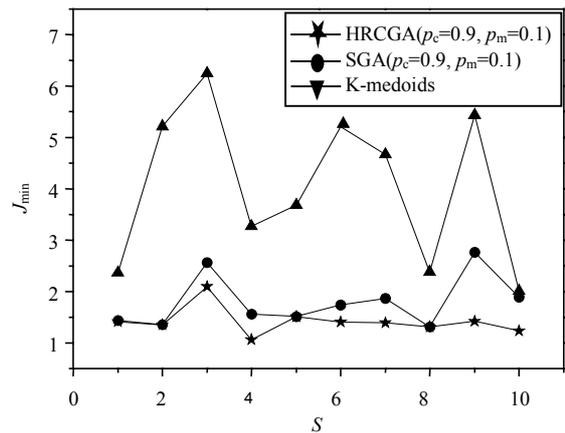


图1 HRCGA对MSN组合选择性能分析

由图1可知, 对于不同的节点样本集, HRCGA对最优MSNs的选择能力明显强于SGA和K-medoids算法。对于K-medoids算法, 达到稳定解时, 其最优目标函数起伏很大。这主要是因为, K-medoids算法存在对初始中心敏感的问题, 并且很容易陷入局部极值, 这与理论分析也一致。与SGA相比, HRCGA通过对SGA的改进, 引入个体进化控制策略, 加强了局部搜索能力, 克服了其早熟收敛问题。

3.2 收敛性分析

对于同一组样本集, 利用HRCGA进行10次实验, 来选择最优MSN组合, 并保证每次HRCGA的初始种群不同。在求解最优MSN组合过程中, 随机选取其中一次实验数据, 其最优个体适应度(optimal fitness)和种群平均适应度(average fitness)进化趋势

如图2所示,其中 F 为个体进化过程中对应的适应度, I 为种群进化迭代次数。

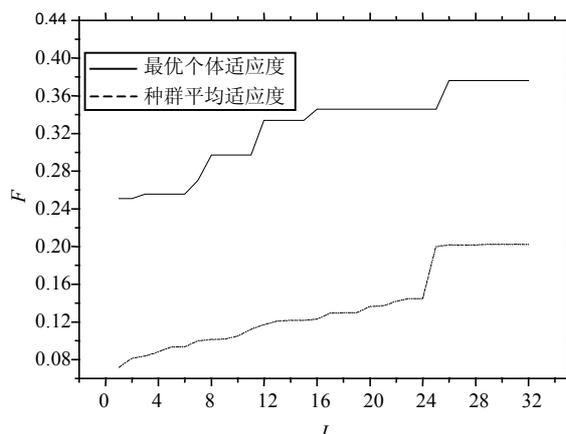


图2 HRCGA收敛性分析

由图2可发现,随着种群进化迭代,其进化方向是向着最优解收敛。

4 结束语

HRCGA通过对SGA的改进,利用实数矩阵编码选择覆盖组播网络MSN组合,既克服了传统K-medoids算法对初始中心敏感和易陷入局部极小值问题,也克服了SGA在全局搜索中的早熟收敛现象。并通过仿真实验验证了HRCGA在覆盖网络组播服务节点选择中的有效性。

参考文献

- [1] YEO C K, LEE B S, ER M H. A survey of application level multicast techniques[J]. Computer Communications, 2004, 27 (15):1547-1568.
- [2] DIOT C, LEVINE B, LYLES J, et al. Deployment issues for the IP multicast service and architecture[J]. IEEE Network, 2000, 14 (1): 78-88.
- [3] BANERJEE S, BHATTACHARJEE B, KOMMAREDDY C. Scalable application layer multicast[C]//ACM SIGCOMM. Pittsburgh: ACM Press, 2002: 205-217.
- [4] FRANCIS P. Yoid: extending the multicast Internet architecture[EB/OL]. [2007-06-02]. <http://www.aciri.org/yoid>.
- [5] ZHANG B C, JAMIN S, ZHANG L X. Host multicast: a framework for delivering multicast to end users [C]//IEEE INFOCOM 2002. New York: IEEE Press, 2002: 1366-1375.
- [6] 沈波, 张宏科, 刘云. 覆盖网络组播压力与深长度的性能评价模型[J]. 系统仿真学报, 2005, 17(5): 1107-1110, 1114.
SHEN Bo, ZHANG Hong-ke, LIU Yun. Performance evaluation models for stress and stretch of overlay network multicast [J]. Chinese Journal of System Simulation, 2005, 17(5): 1107-1110, 1114.
- [7] 郑鹏, 何同林, 刘郁林, 等. 基于实数编码遗传算法的盲信源分离方法[J]. 电子科技大学学报, 2006, 35(3): 295-297, 327.
ZHENG Peng, HE Tong-lin, LIU Yu-lin, et al. Blind source separation method based on real coded genetic algorithm[J]. Journal of University of Electronic Science and Technology of China, 2006, 35(3): 295-297, 327.
- [8] 任子武, 伞冶. 实数遗传算法的改进及性能研究[J]. 电子学报, 2007, 35(2): 269-274.
REN Zi-wu, SAN Zhi. Improvement of real-valued genetic algorithm and performance study[J]. Acta Electronic Sinica, 2007, 35(2): 269-274.
- [9] 行小帅, 潘进, 焦李成. 基于免疫规划的K-means聚类算法[J]. 计算机学报, 2003, 26(5): 605-610.
XING Xiao-shuai, PAN Jin, JIAO Li-cheng. A novel K-means clustering based on the immune programming algorithm[J]. Chinese Journal of Computers, 2003, 26(5): 605-610.
- [10] 陈金山, 韦岗. 遗传+模糊C—均值混合聚类算法[J]. 电子与信息学报, 2004, 24(2): 210-215.
CHEN Jin-shan, WEI Gang. A hybrid clustering algorithm incorporating fuzzy C-Means into canonical genetic algorithm[J]. Chinese Journal of Electronics and Information Technology, 2004, 24(2): 210-215.

编辑 张俊