

求解置换 Flow-shop 调度问题的改进遗传算法

伊华伟¹, 张秋余²

YI Hua-wei¹, ZHANG Qiu-yu²

1. 辽宁工学院 计算机科学与工程学院, 辽宁 锦州 121001

2. 兰州理工大学 计算机与通信学院, 兰州 730050

1. School of Computer Science and Engineering, Liaoning Institute of Technology, Jinzhou, Liaoning 121001, China

2. School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China

YI Hua-wei, ZHANG Qiu-yu. Improved genetic algorithm for permutation Flow-shop scheduling problem. Computer Engineering and Applications, 2007, 43 (22): 41-43.

Abstract: An improved genetic algorithm for permutation Flow-shop scheduling problem is proposed. The algorithm adopts multi-individual-crossover way, takes the threshold value in the process of crossover and mutation, thus expands the search zone of solution space and keeps the variety of the population in the course of optimizing, increases the probability of obtaining optimum solution. Through testing a series of typical Benchmark problems, the results validate the effectiveness of the proposed algorithm.

Key words: genetic algorithm, permutation Flow-shop scheduling problem, multi-individual-crossover, threshold value, population, Benchmark problem

摘要: 提出一种求解置换 Flow-shop 调度问题的改进遗传算法。该算法采用多个体交叉方式, 对交叉过程和变异过程分别进行阈值设置, 实现了在优化过程中扩大解空间的搜索范围和保持种群的多样性, 从而增大了获得最优解的几率。最后对一系列典型的 Benchmark 问题进行仿真测试, 实验结果证实了该改进遗传算法的有效性。

关键词: 遗传算法, 置换 Flow-shop 调度问题, 多个体交叉, 阈值, 种群, Benchmark 问题

文章编号: 1002-8331 (2007) 22-0041-03 **文献标识码:** A **中图分类号:** TP301.6

Flow-shop 调度问题是很多实际流水线生产调度问题的简化模型, 也是一个典型的难度问题, 难以用传统优化方法求解, 因此对其进行研究具有极高的理论价值和实践价值。本文主要对置换 Flow-shop 调度问题进行研究, 它是 Flow-shop 调度问题的一个特例^[1]。

到目前为止, 人们提出了许多求解方法, 如列举法、分支定界、动态规划、NEH 启发式方法等。但是, 由于置换 Flow-shop 调度问题固有的计算复杂性, 对于较大规模问题, 这些优化方法的求解性能常常无法保证。遗传算法自问世以来, 虽然时间很短, 但已显示出了强大的生命力, 在众多领域得到了广泛应用。许多研究表明, 遗传算法对于求解置换 Flow-shop 调度问题的有效求解具有极大的潜力, 它以搜索结果全局近优、快速、易实现等特点成为此类组合优化问题的主流算法, 但由于其易产生早熟现象、局部寻优能力较差等固有特征, 其搜索效率尚有待进一步提高^[2,4]。因此, 针对求解置换 Flow-shop 调度问题提出了一种改进遗传算法。

1 置换 Flow-shop 调度问题描述

一般可以描述为 n 个加工任务 (J_1, J_2, \dots, J_n) 要在 m 台机

器上加工, 每个加工任务需要经过 m 道工序, 每道工序要求不同的机器, n 个加工任务在 m 台机器上的加工顺序相同, n 个加工任务在每台机器上的加工顺序也相同, 加工任务 J_i 在第 j 台机器上的加工时间是给定的, 记为 P_{ij} ($i=1, 2, \dots, n; j=1, 2, \dots, m$)。问题的求解目标是确定 n 个加工任务在每台机器上的最优加工顺序, 使一个或多个性能指标达到最优。对该问题常作如下假设:

- (1) 每个加工任务在机器上的加工顺序是 $1, 2, \dots, m$;
- (2) 每台机器同时只能进行一个加工任务;
- (3) 一个加工任务不能同时在不同的机器上运行;
- (4) 工序不能预定;
- (5) 工序的准备时间与加工顺序无关, 且包含在加工时间中;
- (6) 加工任务在每台机器上的加工顺序相同, 且是确定的;
- (7) 所有的机器都是连续可用的;
- (8) 一个加工任务如果已经在台机器上执行, 不允许中断。

令 t_{ij} 和 $T(i, j)$ 表示加工任务 i 在机器 j 上的加工时间和加工完成时间, 加工顺序为 $\pi = \{i_1, i_2, \dots, i_n\}$, 那么 n 个加工任务

基金项目: 国家科技公关计划项目 (the Key Technologies R&D Program of China under Grant No.2001BA201A32), 国家高技术研究发展计划 (863) (the National High-Tech Research and Development Plan of China under Grant No.2002AA415270)

作者简介: 张秋余 (1966-) 男, 副研究员, 硕士生导师, 主要研究方向: 软件工程、数据库、制造业信息化及信息安全。

m 台机器的置换 Flow-shop 调度问题的完工时间按下列公式计算^[3]：

$$T_{i,j} = t_{i,j} \dots T_{i,k} = \sum_{l=1}^k t_{i,l} \quad (1)$$

$$T_{i,j} = T_{i,j-1} + t_{i,j} \quad (2)$$

$$T_{i,k} = \max\{T_{i-1,k}, T_{i,k-1}\} + t_{i,k} \quad (3)$$

$$i=2, 3, \dots, n, k=2, 3, \dots, m$$

$$C_{max} = T_{n,m} \quad (4)$$

2 基于改进遗传算法的置换 flow-shop 调度的实现

2.1 编码的设计

编码问题是设计遗传算法的首要 and 关键问题。编码技术必须考虑“染色体”的合法性、可行性、有效性以及对问题解空间表征的完全性^[4]。本文采用的是一种实数编码方法，即用加工任务的加工序号排列来表示染色体，按照这样的编码方式，染色体编码可以表示为 $(a_1 a_2 \dots a_n)$ ，其中染色体中基因 a_i 表示第 i 个加工任务的加工序号，其取值范围是 $[1, n]$ 。例如 Car01 Benchmark 问题中有 11 个加工任务，其中一个染色体的编码就可以这样表示 (4 5 7 8 10 6 1 9 11 2 3)。

2.2 适应度函数

使用最大完工时间作为目标函数 $f(x)$ ，具体函数描述见式 (1)~式 (4)。求解的目的是使最大完工时间最小。遗传算法中，通常按照适应度最大来寻优，所以需要目标函数进行改造，本文采用 $\frac{1}{f(x)}$ ，即目标函数的倒数作为适应度函数。

2.3 选择操作

选择方式采用轮盘赌方法，它是一种正比选择策略，能够根据与适应度函数值成正比的概率选出新的种群。按比例适应度分配，是利用比例于各个个体适应度的概率决定其子孙的遗留可能性^[4]。若有 M 个个体待选择，某个个体 i 的适应度为 f_i ，则其被选取的概率为 $P_i = \frac{f_i}{\sum_{i=1}^M f_i}$ ， $i=1, 2, \dots, M$ 。

2.4 交叉操作

2.4.1 交叉方式的设计

采用实数单点交叉，其基本思想就是首先随机选取交叉点，然后选取第一个父个体的交叉点之后的部分基因作为子个体的一部分，再从第二个父个体中按顺序选取合法基因(与选取的第一个双亲的基因不同)填充余下染色体基因，这样可以避免非法个体的产生^[5]。以 Car01 Benchmark 问题为例，交叉操作如图 1 所示。

2.4.2 交叉操作过程

步骤 1 在当前种群中随机选取三个不同的个体 A、B、C；将这三个个体组合成三对不同的父个体对 (A, B)、(A, C)、(B, C)。

步骤 2 对每一对父个体，都重复步骤 3~步骤 6。

步骤 3 设置阈值 R_c (即一定的重复交叉次数)和用来存放产生子个体的数组 p 。

步骤 4 随机选取一个交叉点，对父个体对用 2.4.1 节所设计的交叉方式产生两个子个体，并将产生的子个体放入到 p 中。

步骤 5 若重复交叉次数小于 R_c ，则转步骤 4，否则，转步骤 6。

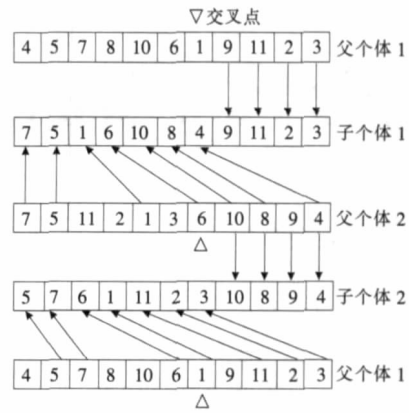


图 1 交叉操作示意图

步骤 6 选出 p 中最好的个体，则将此子个体作为新的个体放入到新的种群。

步骤 7 三对父个体分别进行完以上操作后，就产生了三个新的个体。

步骤 8 重复执行步骤 1~步骤 7，直至产生一个和原种群个体数目相等的新种群。

2.5 变异操作

2.5.1 变异方式的设计

变异算子采用实数互换变异，即随机选取染色体的两个不同基因进行简单交换^[6]。同样以 Car01 Benchmark 问题为例，互换变异操作如图 2 所示。

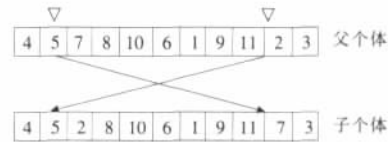


图 2 变异操作示意图

2.5.2 变异操作过程

步骤 1 对由交叉操作产生的新种群中的每一个子个体执行步骤 3~步骤 6。找出当前群体中的最坏个体放入到 S 中。

步骤 2 设置阈值 R_m (即重复交叉次数)。

步骤 3 将要变异的个体放入 B 和 P 。

步骤 4 对 P 中的个体用 2.5.1 节所设计的变异方式产生个体 C 。

步骤 5 如果个体 C 比 B 中的个体好，那么用个体 C 取代 B 中的个体。

步骤 6 如果重复次数小于 R_m ，转到步骤 4，否则，转步骤 7。

步骤 7 如果 B 中个体比原种群中最坏的个体好，那么它将被选择进入新的变异种群，否则，就从原种群中随机选取一个个体作为新的子个体进入到新的种群中。

2.6 最优保留操作

对种群中的个体进行完交叉、变异操作之后，为防止最优解被破坏，进行最优个体保留操作^[7]，将原始种群中的最好个体与当前种群中的最好个体进行比较，若优于当前最好个体，则用其取代当前种群中最好个体，反之，用其取代当前种群中最坏个体。

2.7 改进型遗传算法的特点

(1) 传统遗传算法中，进行交叉操作时，都是选取两个父个体进行遗传操作，本文提出的改进遗传算法采用多个个体的方式^[8]，即选取三个个体，分别进行配对交叉，从而扩大了解空间。

的搜索范围, 并且增加了群体的多样性。

②在交叉操作过程中, 通过设置阈值可以最大限度上继承父代的优良特性, 从而获得更好的子个体, 重复交叉次数达到阈值后, 在产生的子个体群中选出最好的个体, 即使它不如父代种群中的最差个体, 也允许进入新的种群。这样, 在父代种群陷入局部最优解的情况下, 也能跳出局部最优解。

③在变异操作过程中, 同样也设置阈值, 用来获得更好的子个体, 重复变异次数达到阈值后, 将最好的变异个体和父代种群中最坏的个体进行比较, 如果变异个体好, 那么它将被选择进入新种群, 否则, 就从父代种群中随机选取一个个体作为一个新的子个体进入到新的种群中, 为获得最优解提供了条件。

④在每一代对种群进行完选择、交叉和变异操作后, 进行最优个体保留操作, 从而防止最优解的丢失, 加快了算法的收敛速度, 为获得全局最优解提供了保障。

3 仿真实验及结果分析

系统配置为 Pentium IV (1.7 GHz) 128 M, 通过 MATLAB6.5 软件进行仿真实验。将改进遗传算法 (mSGA) 传统遗传算法 (SGA) 对 Car 类和 Rec 类共 10 个典型 Benchmark 问题分别进行测试, 每种算法对每个问题进行 20 次实验, 本文对算法运行参数的确定是依据基本理论, 并经过多次实验, 按照提高算法进化速度来决定算法运行所需的各个参数值: 种群大小 $pop_size=80$; 交叉概率 $p_c=0.85$; 变异概率 $p_m=0.25$; 重复交叉次数 $R_c=60$; 重复变异次数 $R_m=100$; 最大进化代数 $max=80$ 。计算结果如表 1 所示。平均值是每次求得最优结果的和的平均值, BRE 为最优相对误差, 计算公式如下:

$$BRE = \frac{C^W - C^B}{C^B} \times 100\%$$

C^W 为算法求得的最优值, $W \in \{SGA, mSGA\}$, C^B 为理论上的最优值。

表 1 实验结果表

问题	工件数	机器数	C^B	SGA		mSGA	
				BRE	平均值	BRE	平均值
Car01	11	5	7 038	0.00	7 320	0.00	7 046
Car02	13	4	7 166	0.00	7 528	0.00	7 253
Car03	12	5	7 312	1.37	7 595	0.00	7 385
Car04	14	4	8 003	0.00	8 206	0.00	8 021
Car05	10	6	7 720	0.00	7 931	0.00	7 736
Car06	8	9	8 505	0.00	8 682	0.00	8 528
Car07	7	7	6 590	0.00	7 071	0.00	6 961
Car08	8	8	8 366	0.07	8 598	0.00	8 379
Rec11	20	10	1 431	4.47	1 559	0.84	1 468
Rec13	20	15	1 930	4.61	2 057	1.45	1 972
Rec15	20	15	1 950	2.51	2 139	0.77	1 995

实验结果分析如下。

(1)从表 1 可以看出, SGA 求得的结果质量不高, 对于小规模问题 (Car01~Car08) 基本能求得最优解, 但从解的平均值情况来看效果不是很好, 对于问题规模较大的问题 (Rec11, Rec15), 解的平均值和最优相对误差都很不理想, 观察 mSGA 解的情况, 对于小规模问题, 均能求得最优解, 并且平均值接近理论最优值, 对于较大规模问题, 也能取得很好的解。因此, mSGA 的求解性能明显优于 SGA。②图 3 (a) 和图 3 (b) 表明了 SGA 和 mSGA 两种算法对 Car01 和 Rec15 分别进行 20 次反复

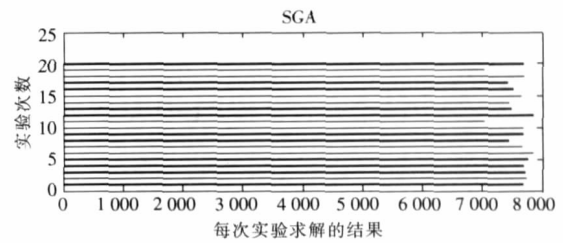


图 3(a) Car01 问题

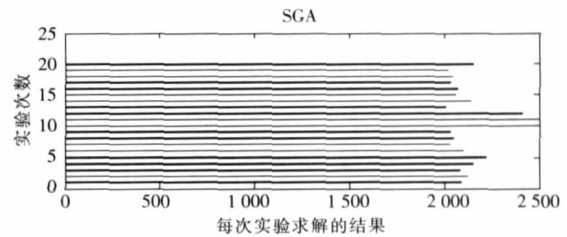


图 3(b) Rec15 问题

实验得到的解的情况, 横轴表示每次实验求得的结果, 纵轴表示实验次数。Car01 问题的理论最优值是 7 038, Rec15 问题的理论最优值是 1 950。由图所示, 对于 Car01 问题, mSGA 在 20 次求解中, 几乎全部达到 7 038, 而 SGA 仅有 2 次达到 7 038; 对于较大规模的 Rec15 问题, mSGA 虽然没有达到理论最优值, 但是解的整体情况比 SGA 明显接近 1 950。因此, 图 3 (a) 和图 3 (b) 也充分说明了 mSGA 的优化性能很高。

4 结束语

针对置换 Flow-shop 调度问题, 本文提出了一种改进遗传算法, 该算法对传统遗传算法从个体交叉数目、交叉过程和变异过程三个方面进行改进, 提高了算法的优化性能, 从而扩大了空间的搜索范围和保持种群的多样性, 为获得全局最优解提供了保障。为了验证改进遗传算法的优化性能, 本文用改进遗传算法对文献[1]中所列的典型 Benchmark 问题进行优化, 并与简单遗传算法进行比较, 实验结果证实了改进遗传算法具有很强的寻优能力。(收稿日期 2007 年 3 月)

参考文献:

- [1] 王凌. 车间调度及其遗传算法[M]. 北京: 清华大学出版社, 2003.
- [2] 玄光男, 程润伟. 遗传算法与工程优化[M]. 北京: 清华大学出版社, 2004.

(下转 82 页)

一种多微粒群协同进化算法

李菲菲¹, 姚坤¹, 刘希玉²

LI Fei-fei¹, YAO Kun¹, LIU Xi-yu²

1. 山东师范大学 信息科学与工程学院, 济南 250014

2. 山东师范大学 管理学院, 济南 250014

1. School of Information Science & Engineering, Shandong Normal University, Ji'nan 250014, China

2. School of Management, Shandong Normal University, Ji'nan 250014, China

E-mail: lff_yk@163.com

LI Fei-fei, YAO Kun, LIU Xi-yu. Multi-particle swarm co-evolution algorithm. Computer Engineering and Applications, 2007, 43 (22): 44-46.

Abstract: Illumined by phenomenon of co-evolution in nature, particle swarm optimization is combined with co-evolution, a multi-particle swarm co-evolution algorithm is presented. In the process of evolution, particle not only exchanges information with other particles in its sub-swarm but also is influenced by other sub-swarms. By doing experiments on three benchmark functions, the results show that the algorithm avoids to trap into local optimum in a certain extent and improves the precision of convergence.

Key words: particle swarm optimization; co-evolution; multi-particle swarm co-evolution

摘要: 受自然界共生现象的启发, 将微粒群算法和协同进化相结合, 提出了一种多微粒群协同进化算法。进化过程中, 粒子不仅要与本子群的其他微粒交换信息, 还要受其他子群体的影响。通过对三个标准函数优化的实验结果表明, 此算法在一定程度上避免了陷入局部极值点并且提高了收敛精度。

关键词: 微粒群算法; 协同进化; 多微粒群协同进化算法

文章编号: 1002-8331(2007)22-0044-03 文献标识码: A 中图分类号: TP301

1 研究背景

微粒群算法最早是在 1995 年由美国社会心理学家 James Kennedy 和电气工程师 Russell Eberhart 共同提出的^[1]。微粒群算法因其在迭代初期的有效性及实现简单等特点在很多领域中得到了广泛地应用, 但是在寻找近似问题的最优解的效率和精度方面还是遇到了很大的困难。这表明, 微粒群算法损失了粒子的多样性, 所有的粒子被迄今为止找到的最好的微粒所吸引。很多研究指出了微粒群算法的缺陷并提出了改进的方法, 比如增加粒子的多样性、引入进化选择机制以及速率更新方程中的可调参数。

受协同进化遗传算法的启发, Shi 和 Krohling^[2]提出了一种基于两个 PSO 协同进化的求解极大极小值的算法。Asmara, Krohling 和 Hoffmann^[3]在此算法的基础上又提出了一种“accelerated co-evolutionary PSO”, 提高了算法的收敛速度并将此算法应用到机器人臂计算机扭矩控制中。Ashraf M. Abdelbar, Sherif Ragab 和 Sara Mitri^[4]提出了将协同进化微粒群算法应用到古埃及的一种对弈的棋盘游戏中。Mohammed El-Abd 和 Mohamed Kamel^[5]在他们的论文中指出了两个相互合作的微粒群之间如何进行信息交换。J.J. Liang 和 P.N. Suganthan^[6]提出了

一种动态多微粒群优化算法。Krohling 和 Hoffmann 等^[7]提出了用协同进化微粒群算法来解决带约束的极大极小值问题。在国内, 很多学者也对协同进化微粒群算法进行了研究。Ben Niu, Yunlong Zhu 和 Xiaoxian He^[8,9]提出了多群体合作微粒群算法, 并将其用在了为动态系统构建模糊模式上。李爱国^[10]提出了一种两层结构的多粒子群协同优化算法, 底层用多个粒子群相互独立地搜索解空间以扩大搜索范围, 上层用 1 个粒子群追逐当前全局最优解以加快算法收敛。Hong-yun Meng, Xiao-hua Zhang 和 San-yang Liu^[11]提出将协同进化微粒群算法用到求解多目标优化问题中, 此算法针对多目标问题设计了协同进化算子、竞争变异算子和新的选择机制, 通过共享和交换群体间的信息从而收缩搜索区域, 维持了种群的多样性, 避免了陷入最优。王俊年、申群太等^[12,13]提出了将多种群协同进化微粒群算法应用到神经网络自适应噪声消除系统和径向神经网络设计中。

虽然协同进化微粒群算法在一定程度上得到了发展, 但是在多种群协同的机制上有待于改善, 在微粒群算法和遗传算法的结合上还有待于进一步研究。本文在已有文献的基础上提出了一种多微粒群协同进化算法, 实验表明此算法在一定程度上

基金项目: 国家自然科学基金 (the National Natural Science Foundation of China under Grant No.6037405); “泰山学者”建设工程专项经费。

作者简介: 李菲菲 (1981-), 硕士研究生, 研究方向为群体智能、数字图像处理; 姚坤 (1981-), 硕士研究生, 研究方向为群体智能、创新概念设计; 刘希玉 (1964-), 博士, 教授, 博士生导师, “泰山学者”, 研究方向为神经网络与进化计算及其应用等。

避免了陷入局部极值点并且提高了收敛精度。

2 微粒群算法

设 D 在维搜索空间中, 有 m 个没有重量和体积的粒子组成的一个种群, 这些粒子在搜索空间中以一定的速度飞行, 飞行速度由个体的飞行经验和群体的飞行经验进行动态调整。其中第 i 个粒子的位置为 $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$, 飞行速度为 $V_i = (v_{i1}, v_{i2}, \dots, v_{in})$, $P_i = (p_{i1}, p_{i2}, \dots, p_{in})$ 为粒子 i 所经历的最好位置, 也就是微粒 i 所经历过的具有最好适应值的位置, 称为个体最好位置。 $P_g = (p_{g1}, p_{g2}, \dots, p_{gn})$ 代表整个微粒群找到的最优位置。将 X_i 代入要优化的问题中计算适应值, 粒子状态更新方程为:

$$v_{ij}(t+1) = wv_{ij}(t) + c_1r_{1j}(t)[p_{ij}(t) - x_{ij}(t)] + c_2r_{2j}(t)[p_{gj}(t) - x_{ij}(t)] \quad (1)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad (2)$$

式中 w 称为惯性权重, $j=1, 2, \dots, m$ 表示粒子个数, $j=1, 2, \dots, D$ 表示搜索空间维数, c_1, c_2 为学习系数, r_1, r_2 是之间的随机数。

3 多微粒群协同进化算法

多微粒群协同进化 (MPSC) 算法采用一个 master 种群和若干个 slave 种群, 每个 slave 种群独立进化, 每次迭代结束后将最好的微粒复制到 master 种群, 然后 master 种群开始进化, 找出最好的微粒, 将此微粒信息反馈给各 slave 种群, 然后各 slave 种群据此信息修改微粒的速度继续进化。各 slave 种群通过 master 种群进行信息交流, 各种群信息交流过程如图 1 所示。各 slave 种群进化过程采用文献[14]中的 GA 和 PSO 混合算法-GSO 算法, GSO 算法的流程如图 2 所示。

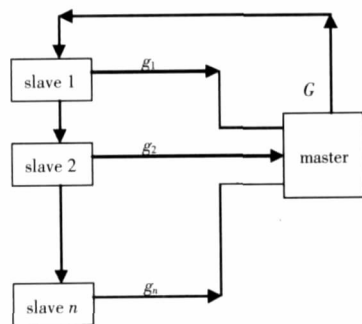


图 1 多微粒群协同进化算法信息交流示意图

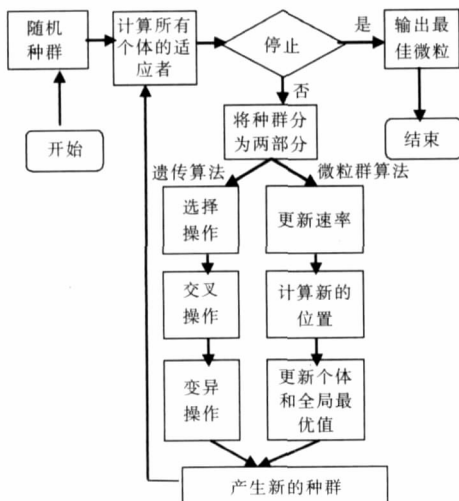


图 2 GSO 算法流程图

对 GSO 算法作一些改进, 微粒的速率和位置更新公式为:

$$v_j^i(t+1) = \chi \times (w \times v_j^i(t) + r_1 c_1 (p_j^i - x_j^i(t)) + r_2 c_2 (q_j - x_j^i(t)) + r_3 c_3 (G - x_j^i(t))) \quad (3)$$

$$x_j^i(t+1) = x_j^i(t) + v_j^i(t+1) \quad (4)$$

$v_j^i(t+1)$ 为第 i 个 slave 种群的第 j 个微粒在第 $t+1$ 代的速度;

$x_j^i(t)$ 为第 i 个 slave 种群的第 j 个微粒在第 t 代的位置;

w 为惯性权重。随着进化过程, 根据如下公式调整惯性权重:

$$w = w_{\max} - \frac{w_{\max} - w_{\min}}{\text{iter}_{\max}} \times \text{iter} \quad (5)$$

$$\chi = \frac{2}{2 - \varphi - \sqrt{\varphi^2 - 4\varphi}} \quad \varphi = c_1 + c_2, \varphi > 4 \quad (6)$$

χ 为收缩因子;

r_1, r_2, r_3 为在 $[0, 1]$ 之间的随机数;

c_1, c_2, c_3 为学习系数。

p_j^i 为第 i 个 slave 种群的第 j 个微粒的个体历史最好位置;

q 为第 i 个 slave 种群的全局最好位置;

G 为 master 种群的全局最好位置。

交叉操作: 采用单点交叉, 将每个粒子的当前位置从维数的中间交换重组, 例如粒子的维数为 10, 粒子一 6 维-10 维的值等于粒子二 6 维-10 维的值, 粒子二 1 维-5 维的值等于粒子一 1 维-5 维的值, 其余各项值保持不变。

变异操作: 随机生成变异粒子的当前位置, 变异粒子其余的值保持不变。

master 种群采用 PSO 算法, 其速度和位置更新公式为:

$$v_i^M(t) = \chi \times (w \times v_i^M(t) + c_1 r_1 (p_i^M - x_i^M(t)) + c_2 r_2 (G - x_i^M(t))) \quad (7)$$

$$x_i^M(t) = x_i^M(t) + v_i^M(t) \quad (8)$$

$v_i^M(t)$ 为 master 种群中微粒 i 更新后的速度;

$x_i^M(t)$ 为 master 种群中微粒 i 更新后的位置;

q^M 为当前 master 种群中最好微粒的位置;

G 为上次迭代结束后 master 种群中最好微粒的位置。

master 种群的速度和位置更新结束后, 要计算每个微粒的适应值, 找出最好的微粒。如果最好的微粒的位置差于 G , 则继续保留 G 。否则, 将 G 更新为最好微粒的位置。

多微粒群协同进化算法流程

Begin

初始化每个 slave 群体中的每个微粒;

for $i=1$ to n

计算每个微粒的适应值找出最好的微粒,

将最好的微粒复制到 master 种群;

endfor

对 master 种群中的微粒找出最好的微粒 G ;

while (G 不满足条件或未达到迭代次数)

for $i=1$ to n

每个 slave 种群按概率分为两部分;

其中一部分进行交叉、变异操作;

另一部分按照公式 (3) 和 (4) 更新速度和位置;

计算每个微粒的适应值找出最好的微粒;

更新 p_i 和 q ;

将最好的微粒复制到 master 种群;

endfor

表2 MPSC算法运行20次的平均最好适应值对比(平均值±标准差)

Functions	PSO	FPSO	HPSO	PSCO	MPSC
$f_1(x)$	317.78 ±55.16	190.70 ±57.33	176.83 ±40.66	100.60 ±54.47	7.466 8 ±1.3195 3
$f_2(x)$	24.92 ±4.90	22.33 ±2.71	11.99 ±1.58	10.635 8 ±2.58	3.376 2 ±2.415 2
$f_3(x)$	0.21 ±0.044	0.025 7 ±0.218	0.047 ±0.023	0.007 ±0.014	0.005 3 ±0.011 5

对 master 种群中的每个微粒根据公式(7)
和(8)更新速度和位置,计算适应值更新G;

endwhile

End

4 仿真实验

本文比较 MPSC 算法与模糊 PSO (FPSO)、杂交 PSO (HPSO) 以及 PSCO 等算法^[10]对 Rosenbrock 函数 (f_1)、Rastrigrin 函数 (f_2) 和 Griewank 函数 (f_3) 等3种基准函数的优化精度。

Rosenbrock 函数又称为香蕉函数,其最优解为 $x^* = (1, 1, \dots, 1)$ 极小值为 $f(x^*) = 0$ 。

$$f_1(x) = \sum_{i=1}^{n-1} (100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2), \quad -100 \leq x_i \leq 100$$

Rastrigrin 函数是一个多峰函数,有很多正弦凸起的局部极小点,它的全局极小点在 $x^* = (0, 0, \dots, 0)$ 处,全局极小值 $f(x^*) = 0$ 。

$$f_2(x) = \sum_{i=1}^n ((x_i^2 - 10 \cos(2\pi x_i)) + 10), \quad -10 \leq x_i \leq 10$$

Griewank 函数也是一个多峰函数,它的全局极小点在 $x^* = (0, 0, \dots, 0)$ 处取得,全局极小值 $f(x^*) = 0$ 。

$$f_3(x) = \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1, \quad -600 \leq x_i \leq 600$$

MPSC 算法用 .NET2003 实现,进行实验时,每个微粒设为 10 维,取微粒群子群体数 $S=6$;每子群中粒子数 $m=20$,相应主群体中粒子数 $m=6$,MPSC 算法的每次迭代过程中每个子群中的粒子按照概率 0.5 被分成两个群体:一个是 PSO 群体按照公式(3)和(4)更新速度和位置进行 PSO 操作,另一个是 GA 群体进行 GA 操作;主群体进行 PSO 操作时按照公式(7)和(8)更新粒子速度和位置,其惯性权重也是随迭代次数从 0.7 递减至 0.3,学习系数 c_{master} 的值和子群体的一样;具体参数设置请参看表 1。实验经典 PSO 算法时,取 $m=20$, $c_1=c_2=2.0$, $w=0.7$, $V_{\text{max}}=100$ (对于 $f_1(x)$), $V_{\text{max}}=10$ (对于 $f_2(x)$), $V_{\text{max}}=600$ (对于 $f_3(x)$)。

表1 MPSC算法优化标准函数的参数设置

最大速度	w	c_1	c_2	c_3	c_{master}	交叉变异		迭代次数	
						概率	概率		
$f_1(x)$	20	0.7 递减至 0.3	2.05	2.05	2.05	2.05	0.5	0.1	10 000
$f_2(x)$	2	0.7 递减至 0.3	2.05	2.05	2.05	2.05	0.2	0.1	10 000
$f_3(x)$	200	0.7 递减至 0.3	2.05	2.05	2.05	2.05	0.7	0.1	10 000

MPSC 算法运行 20 次取平均最好适应值与其他算法进行比较,如表 2 所示实验结果,MPSC 算法相对于前几种算法显示了更好的优化性能。特别对于函数 $f_2(x)$,当 MPSC 算法的迭代次数在 30 000 次以上时,每次运行基本都能得到全局最优值。

5 结论与展望

本文在已有文献的基础上将微粒群算法与协同进化算法结合,集中两个算法的优点,提出了一种多微粒群协同进化算法,通过对三个标准函数优化的结果表明该算法在一定程度上加快了收敛速度、避免了陷入局部极值点的问题。下一步要做

的工作有:

(1) 因为此算法的交叉操作的交叉点是固定的,是否交叉点随机产生会产生更好的结果还有待于进一步研究。

(2) 因为此算法的参数较多,各项参数的不同设置会对运行结果产生什么影响,会不会产生更好的结果还有待于进一步研究。(收稿日期:2007年3月)

参考文献:

- [1] Kennedy J, Eberhart R C. Particle swarm optimization[C]//Proc IEEE Int Conf on Neural Networks IV. Piscataway, NJ: IEEE Service Center, 1995: 1942-1948.
- [2] Shi Y, Krohling R. Co-evolutionary particle swarm optimization to solving min-max problems [C]//Proc IEEE Congress on Evolutionary Computation, Honolulu, Hawaii, May 2002: 1682-1687.
- [3] Asmara A, Krohling R A, Hoffmann F. Parameter tuning of a computed-torque controller for a 5 degree of freedom robot arm using co-evolutionary particle swarm optimization[C]//Proceedings of 2005 IEEE Conference on Swarm Intelligence Symposium 2005: 162-168.
- [4] Abdelbar A M, Ragab S, Mitri S. Applying co-evolutionary particle swarm optimization to the egyptian board game seega[C]//Proceedings of 2004 IEEE International Joint Conference on Neural Networks 2004: 244-248.
- [5] Mohammed El-Abd, Mohamed Kamel. Information exchange in multiple cooperating swarms[C]//Proceedings of the 2005 Conference on Genetic and Evolutionary Computation 2005: 269-270.
- [6] Liang J J, Suganthan P N. Dynamic multi-swarm particle swarm optimizer[C]//Proceedings of 2005 IEEE Conference on Swarm Intelligence Symposium 2005: 124-129.
- [7] Krohling R A, Hoffmann F, Leandro dos, et al. Co-evolutionary particle swarm optimization for min-max problems using Gaussian distribution[C]//Proceedings of CEC2004 Congress on Evolutionary Computation 2004: 959-964.
- [8] Niu Ben, Zhu Yun-long, He Xiao-xian. Multi-population cooperative particle swarm optimization[C]//LNAI 3630: ECAL 2005, 2005: 874-883.
- [9] Niu Ben, Zhu Yun-long, He Xiao-xian. Construction of fuzzy models for dynamic systems using multi-population cooperative particle swarm optimizer[C]//LNAI 3613: FSKD 2005, 2005: 987-1000.
- [10] 李爱国. 多粒子群协同优化算法[J]. 复旦学报: 自然科学版, 2004, 43(5): 923-925.
- [11] Meng Hong-yun, Zhang Xiao-hua, Liu San-yang. A co-evolutionary particle swarm optimization-based method for multiobjective optimization[C]//LNAI 3809: AI 2005, 2005: 349-359.
- [12] 王俊年, 申群太, 沈洪远, 等. 基于多种群协同进化微粒群算法的径向基神经网络设计[J]. 控制理论与应用, 2006, 23(2): 251-255.
- [13] 王俊年, 申群太, 沈洪远, 等. 基于协同进化微粒群算法的神经网络自适应噪声消除系统[J]. 计算机工程与应用, 2005, 41(13): 20-23.
- [14] Grimaldi E A, Gandelli A, Grimaldi F, et al. A new hybrid technique for the optimization of large-domain electromagnetic problems[C]//Proceedings of URSI GA 2005, 2005.

一个基于混沌映射的图像加密并行算法

周庆, 胡月, 廖晓峰

ZHOU Qing, HU Yue, LIAO Xiao-feng

重庆大学 计算机学院, 重庆 400044

College of Computer Science, Chongqing University, Chongqing 400044, China

ZHOU Qing, HU Yue, LIAO Xiao-feng. Parallel image encryption algorithm based on chaos map. Computer Engineering and Applications, 2007, 43 (22): 47-49.

Abstract: Presents some requirements for parallel image encryption algorithm. Proposes a parallel image encryption algorithm based on chaos map. Both theoretic analysis and computer simulation show this new algorithm is secure and suitable for parallel computation.

Key words: image encryption, parallel computing, chaos map

摘要: 研究了图像加密并行算法应该满足的要求, 并提出了一个基于混沌映射的并行算法。理论分析和实验结果表明该算法适宜于并行计算环境。

关键词: 图像加密; 并行计算; 混沌映射

文章编号: 1002-8331(2007)22-0047-03 文献标识码: A 中图分类号: TP301

1 图像加密算法的并行化

多媒体技术和 Internet 技术的发展使数字图像的应用日益丰富。如今, 数字图像被广泛应用于娱乐、商业、教育、医疗和军事等领域。在这些领域中, 涉及敏感内容的图像都需要进行加密, 以确保其内容不被非授权者窃听。

加密图像最直接的方法是使用经典的加密算法如 AES 对图像文件进行加密。但是由于图像与文本具有不同特性, 这种方法并不能完全保证图像的安全性。对某些特殊的图像, 特别是那些冗余度较大的图像, 直接使用传统加密方法的加密效果往往不能令人满意 (参见图 1)。另一方面, 采用传统加密方法没有充分利用图像的一些性质, 而图像通常具有较大的数据量, 这使得图像的加密速度较慢。为了克服以上两个缺点, 许多快速图像加密算法被提出^[1-12], 其中基于混沌映射的图像加密算法引起了研究者的广泛兴趣^[1-7]。这一类加密算法通常借助了混沌映射的许多特性:

(1) 随机性: 混沌系统能够以确定性的方式产生长周期的

伪随机序列。

(2) 敏感性: 混沌系统对状态初值和系统敏感, 即所谓的“蝴蝶效应”。

(3) 简单性: 混沌系统通常以简单的运算产生复杂的行为。

(4) 遍历性: 混沌系统能等概率地经过所有的状态。

除以上特性以外, 一些二维混沌映射非常适合用于图像像素的置换, 因此混沌映射被广泛用于数字图像加密的研究中。

尽管使用混沌映射的方式多种多样, 但归纳起来基于混沌映射的图像加密算法^[1-7]都运行在同一框架下:

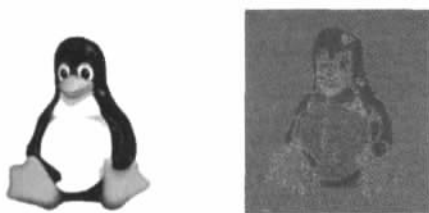
步骤 1 使用混沌映射对像素进行转换。

步骤 2 使用类似 CBC (密码分组链接) 的模式加密每个像素。

步骤 3 重复步骤 1 和步骤 2 直到加密结束。

在步骤 2 中, 由于使用了类似 CBC 的模式, 即当前像素的加密要受到前一像素的影响, 从而有效地实现了扩散。但是类 CBC 模式的引入也使得加密并行化变得不可行。由于只有当前一个像素加密完成后才能加密当前像素, 从而使整个加密变为一个串行过程。

目前并行计算已成为一种发展趋势, 无论是采用多核技术的通用 PC 机, 还是使用 FPGA/CPLD 的专业芯片都支持并行计算方式。采用并行计算可以大大提高运算速度, 这对数据量较大的图像加密应用是非常有利的, 因此研究图像加密的并行算法具有一定的实用价值。本文的结构安排如下: 第 2 章首先给出图像加密并行算法应该满足的要求, 第 3 章介绍两种基本



(a) 明文 (b) 密文
图 1 使用 AES 的图像加密

基金项目: 国家自然科学基金 (the National Natural Science Foundation of China under Grant No.60573047), 重庆市自然科学基金 (the Natural Science Foundation of Chongqing City of China under Grant No.8509, No.8986-3), 新世纪优秀人才支持计划, 重庆大学研究生创新基金 (No.200609Y1B0150174)。

的加密技术 AES (Advanced Encryption Standard) 算法和 Kolmogorov 映射,第 4 章首先给出一个完整的图像加密并行算法,然后对其并行特性进行了分析,在第 5 章中,给出实验及其分析,以作为对加密算法的安全性检验,第 6 章总结全文。

2 图像加密并行算法及其要求

在本文所指的并行模式中,多个 PE (Processing Elements) 同时加密一个图像,每个 PE 拥有私有的内存,各个 PE 之间存在数据通信。图 2 给出了 3 个 PE 实现并行加密的示例。

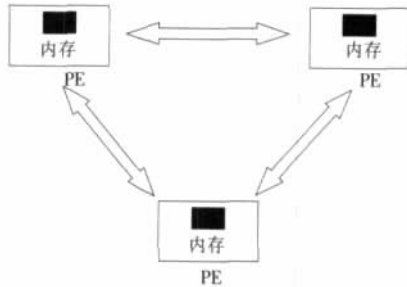


图 2 图像加密的并行模式

一个好的图像加密并行算法应该满足以下的要求:

(1) 完全扩散

扩散指明文的微小变化能够引起密文较大的改变,它是安全性的一个基本要求。在图 1 中,加密失败的一个重要原因就是各个分组的加密是独立的,没有引入扩散。为了解决这一问题,现有的混沌加密通过类似 CBC 的加密模式来实现像素之间的扩散,从而提高了其安全性。为了实现加密的并行化,应该舍弃 CBC 模式,但仍然要实现扩散的要求。更进一步,为了最大化安全,要求当一个像素改变时,要求所有的像素都要发生改变,称为完全扩散。

(2) 通信负载均衡

为了实现扩散的要求,各个 PE 之间必然要进行通信。由于在一轮加密中,通信时间取决于单个 PE 最长的通信时间,因此一个好的并行算法应该使各个 PE 之间的通信量非常接近,即实现通信负载均衡。

(3) 计算负载均衡

同样道理,希望各个 PE 的计算量足够接近,称为计算负载均衡。

(4) 临界区管理

在并行计算中,如果多个 PE 可能同时读或写同一存储区,则该存储区称为临界区。为了防止不期望的错误,一个好的并行算法应该实现有效的临界区管理。

3 AES 算法与 Kolmogorov 映射

Rijdael 是由两位比利时学者 Joan Daemen 和 Vincent Rijmen 提出的加密算法。2000 年 10 月美国国家标准与技术研究所 (NIST) 正式确定该算法作为 AES,以取代逐渐过时的 DES (Data Encryption Standard) 算法。与 DES 相比, AES 的速度更快,并且在常规条件下可以抵御目前已知的任何攻击^[13]。AES 是一个多轮加密体制,每一轮包含密钥加法、字节代替、行移位和列混合 4 个操作。关于 AES 算法更详细的内容可参见文献^[14]。

离散 Kolmogorov 映射最先是由 Pichler 和 Scharinger 引入到图像加密中的^[1],之后 Fridrich 在此基础上进行了改进^[2]。

离散的 Kolmogorov 映射可以用式 (1) 表示:

$$T_{n_s}(x, y) = (q_s(x - F_s) + (y \bmod q_s), F_s + (y \operatorname{div} q_s)) \quad (1)$$

其中 $n_s = (n_1, n_2, \dots, n_k)$, n_s 为正整数, $p_s = N/n_s$, $\sum_{s=1}^k n_s = N$, N 为图像的边长,并且任意的 n_s 整除 N 。 F_s 可以由式 (1) 定义:

$$F_s = \begin{cases} 0 & s=1 \\ n_1+n_2+\dots+n_{s-1} & s=2, 3, \dots, k-1 \end{cases} \quad (2)$$

直观地,一个离散的 Kolmogorov 映射可以用图 3 几何表示。在图 3 中, $N \times N$ 的图像首先按照 n_s 划分成多个 $n_s \times N$ 矩形。每个矩形被进一步等分成 n_s 个 $n_s \times p_s$ 的小矩形。进行 Kolmogorov 变换后,每个小矩形的像素就被映射到同一行中。经过多次变换后,图像原有的结构将被彻底破坏,从而呈现出一种杂乱无章的加密图像。

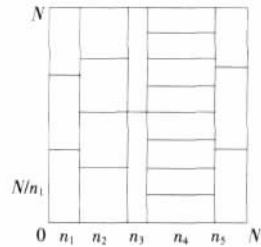


图 3 Kolmogorov 映射的几何表示

Scharinger 已证明经过 $4 \lg N$ 轮 Kolmogorov 变换后,图像密文具有如下的性质^[15]:

(1) 遍历性 (Ergodicity): 任意初始像素将等概率地移动到任何其它位置。

(2) 指数发散 (Exponential divergence): 相邻像素的位置将按指数进行发散。

(3) 混合特性 (Mixing property): 任何子区域内的像素将均匀地分布到整个区域。

图 4 显示了图像 'Lena' 在经过 1 轮和 9 轮 Kolmogorov 变换后的状态 ($= [128 \ 256 \ 128]$)。



(a)明文 (b)1 轮变换后的图像 (c)9 轮变换后的图像

图 4 Kolmogorov 变换

4 一个图像加密并行算法

设 $N \times N$ 的图像由 n 个 PE 并行加密,且 n 整除 N (如不整除可补足像素使 n 整除 N) 则加密的具体过程如下:

步骤 1 任务分配:第 i ($i=0, 2, \dots, n-1$) 个 PE 负责 i 的模 n 同余类所在行的加密。例如对 $n=4$,第 2 个 PE 负责第 2, 6, 10, ... $N-n+2$ 等行的加密。

步骤 2 每个 PE 对其负责的所有行上的像素进行 Kolmogorov 变换,其中

$$= (n, n, \dots, n)_{q_s} \quad (3)$$

其中 $q_s = N/n$ 。注意,变换后的像素可能不再由该 PE 负责加密,这时将涉及 PE 间的数据通信。

步骤 3 每个 PE 对其负责的每一行上的像素分别用 CBC 模式进行 AES 加密。

步骤4 继续步骤2直到加密完成。

该算法能够满足并行加密算法的4个要求:

(1) 完全扩散要求

假设AES有足够好的扩散效果,经过第1轮CBC模式的AES加密后,任意像素细微的变化将扩散到该像素所在行上的 N 个像素。若按照式(3)进行设定,则经过第2轮的Kolmogorov变换后,容易证明这 N 个像素将分散到 q 个不同的行。因此经过第2轮的AES加密后,扩散效果扩大到 q 行上的全部像素。而第3轮加密后,扩散效果扩大到整个图像。

(2) 通信负载平衡

假设按照式(3)进行设定,同样容易证明对每个PE,交换的数据量均为 $N \times N \times (q-1)q^2$,因此本算法可以实现各PE之间的通信负载平衡。

(3) 计算负载平衡

显而易见,每个PE加密的数据量均为 q 行像素,因此该算法实现了各PE之间的计算负载平衡。

(4) 临界区管理

由于在整个计算过程中,不会出现两个PE同时读或写同一像素,因此该算法不存在临界区问题,因此不需要进行临界区管理。

5 实验及分析

为了验证算法的安全性,本章给出了一个实验结果,并对此进行了分析。图像明文是 256×256 的灰度图像'Lena'(图5(a))。在本实验中,PE数 $n=4$,加密轮数等于9,AES密钥为'0123456789ABCDEF',图5(b)显示了图像加密后的结果。部分实验结果如下。

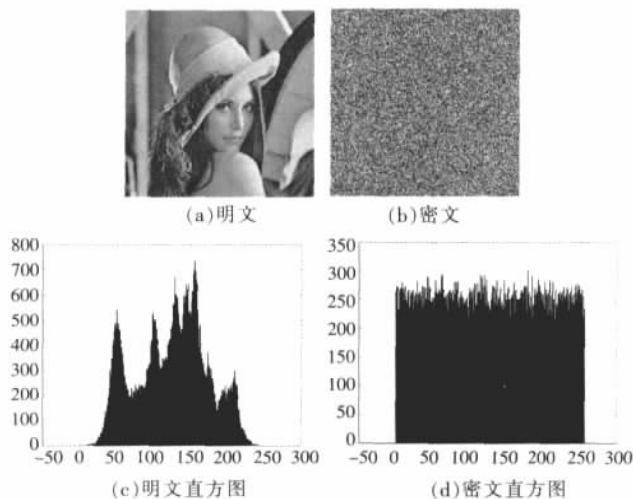


图5 明文和密文图像及其直方图

(1) 直方图

图5(c)和图5(d)显示了加密前后图像的直方图。从图中可以看出加密后图像的直方图非常地均匀,从而提高了图像的

安全性。

(2) 相邻像素的相关系数

图像的一个显著特征是相邻像素的相关性很高,一个成功的图像加密算法应该去除这种相关性。分别从图像的水平、竖直和对角方向随机地选取1000对像素,并对比它们加密前后的相关系数,对比结果显示在表1中。

表1 明文和密文图像中相邻像素的相关系数

	明文	密文
水平	0.983 1	0.018 9
竖直	0.957 2	-0.029 5
对角	0.943 5	0.028 7

(3) NPCR 分析

NPCR (Number of Pixel Change Rate)指当明文的一个像素改变时,密文像素的改变率。表2给出了当明文的最后一个像素的最后一位改变时,加密图像在不同轮数中的NPCR值。注意,由于每个像素有256个可能的取值,像素值在加密前后可能保持不变,其概率等于 $1/256$ 。由表2可以看出,在第2轮加密之后,扩散的范围约为 $1/n=1/4$ 。而3轮加密后实现了完全扩散。

(4) UACI 分析

UACI (Unified Average Changing Intensity)指当明文改变一个像素时,密文之间差的绝对值。同样改变明文最后一个像素的最后一位,其UACI除以最大值255后的结果在表3中给出。注意,对于在 $[0,1]$ 之间均匀分布的两个随机向量,其UACI值应等于 $1/3$ 。

(5) 其它安全问题

本文提出的算法中,Kolmogorov变换是公开的,并不受到密钥的控制,这一点与其它基于混沌映射的图像加密算法有很大的不同。尽管如此,该算法的安全性并不会降低。因为Kolmogorov映射在算法中的主要功能是在不影响算法并行性的前提下将AES产生的局部扩散扩大到整个图像,算法的安全性仍然依赖于AES加密算法。事实上,现有的经典加密算法,如DES或AES采用的置换操作均是公开的,仅有异或运算才需要引入密钥。因此采用公开的Kolmogorov变换并不会降低算法的安全性。相反地,如果使用密钥来控制Kolmogorov变换则会增加算法的计算量,同时不恰当的密钥还会影响算法的安全性和并行度。

6 结论及进一步工作

本文首先分析了现有的各种图像加密算法,指出它们在实现并行计算时的缺陷。针对这一问题,首先提出图像加密并行算法应该满足的要求。然后提出了一个基于AES算法和Kolmogorov映射的图像加密并行算法。该算法能够满足图像加密并行算法需要的4个要求,实验结果也验证了其安全性。在下一步工作中,将研究AES的替代算法,以期利用图像本身的性

表2 各轮加密后的NPCR值

轮数	1	2	3	4	5	6	7	8	9
NPCR	0.003 8	0.249 9	0.995 9	0.996 4	0.996 1	0.996 2	0.995 9	0.996 2	0.996 1

表3 各轮加密后的UACI值

轮数	1	2	3	4	5	6	7	8	9
UACI	0.001 3	0.061 1	0.334 9	0.333 3	0.333 9	0.333 6	0.333 3	0.333 6	0.333 4

(下转 66 页)

基于阻塞调度的优先级反转解决策略

陈尹立¹, 彭诗力¹, 廖春蓝²

CHEN Yin-li¹, PENG Shi-li¹, LIAO Chun-lan²

1. 广东金融学院, 广州 510520

2. 番禺职业技术学院, 广州 511483

1. Guangdong University of Finance, Guangzhou 510520, China

2. Panyu Polytechnic, Guangzhou 511483, China

E-mail: pengshipower@163.com

CHEN Yin-li, PENG Shi-li, LIAO Chun-lan. Priority inversion solution based on blocking scheduling. Computer Engineering and Applications, 2007, 43 (22): 50-52.

Abstract: At present, an approach solving solution is priority ceiling protocol ordinarily, but the resources efficiency of the system based on the protocol is low. To the question, the paper analyses the necessary condition of priority inversion and brings forward a new blocking scheduling algorithm. It can effectively prevent priority inversion and detect deadlocks, and has higher resources efficiency and broader applicability.

Key words: blocking scheduling, priority inversion, priority ceiling protocol, deadlock

摘要: 目前通常采用优先级上限协议解决策略, 但该协议的系统资源利用率较低。针对该问题, 深入分析了优先级反转的必要条件并提出了一种新的阻塞调度算法, 能有效的避免优先级反转和检测死锁现象, 同时具有较高的资源利用率和较广的适用范围。

关键词: 阻塞调度, 优先级反转, 优先级上限协议, 死锁

文章编号: 1002-8331(2007)22-0050-03 文献标识码: A 中图分类号: TP301

1 引言

优先级反转是指在优先级驱动的可剥夺内核中高优先级任务被低优先级任务阻塞, 其阻塞时间无法预测的现象, 严重恶化了系统的实时性和可调度性^[1]。目前实时操作系统 (RTOS) 一般采用优先级继承协议和优先级上限协议来避免优先级反转, 但两者都是次优协议^[1,3]。

2 优先级反转现象及其原因

假设有 3 个任务 T_1 、 T_2 、 T_3 , 优先级由高到低, T_1 和 T_3 共享临界资源 S 。 T_3 先就绪, 并占用临界资源 S , 在 t_1 时刻 T_1 就绪并抢占 T_3 运行, 此时 T_3 并没有释放资源 S 。 t_2 时刻, T_1 因访问临界资源 S 而被堵塞, T_3 继续执行, 以上过程是正常的任务切换。在 t_3 时刻, T_2 就绪并抢占 T_3 运行, 此时 T_3 仍然没有释放资源 S 。 此时 T_1 和 T_3 都无法运行, T_1 被临界资源堵塞, T_3 被 T_2 抢占。只有等 T_2 执行完毕, T_3 接着执行并释放资源 S , T_1 才能继续执行。 T_1 和 T_2 之间并没有同步关系, 但是 T_2 先于 T_1 执行完毕。正常情况下, T_1 因堵塞而等待时间为 T_3 占用资源 S 的时间, 是可以预测的。如果发生优先级反转则 T_1 的等待时间包括 T_2 的执行时间, 无法预测阻塞时间, 优先级反转示意图见图 1。

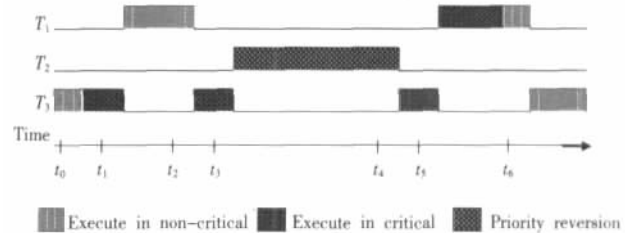


图 1 优先级反转示意图

优先级反转现象严重影响了实时系统的时间正确性, 其发生的必要条件可以归结为 3 点:

(1) 共享临界资源的各个任务优先级不相等或不连续, 存在中间优先级任务。

(2) 需要临界资源的任务同时受到临界资源和优先级的约束, 其中临界资源约束高于优先级约束, 且两者的约束对象不一致。

(3) 低优先级任务占用临界资源并阻塞高优先级任务期间, 允许被中间优先级任务抢占。

其中, 条件 (1) 可以在分配任务优先级时, 慎重考虑, 尽量避免。优先级上限协议和级继承协议主要是针对条件 (2), 提高占用临界资源任务的优先级, 尽快释放临界资源。阻塞

基金项目: 国家自然科学基金 (the National Natural Science Foundation of China under Grant No.50275152)。

作者简介: 陈尹立 (1964-) 男, 博士, 副教授, 研究方向为操作系统及软件应用; 彭诗力 (1977-) 男, 硕士, 系统分析师, 研究方向为嵌入式系统; 廖春蓝 (1979-) 女, 硕士, 研究方向为单片机和 ARM 技术。

调度算法是针对条件(3)提出来的,在低优先级任务因占用临界资源而阻塞高优先级任务期间不允许被中间优先级任务抢占。

3 经典解决策略及其缺陷分析

优先级继承协议的基本思想是:当任务阻塞高优先级任务时,将其优先级暂时提高到被它阻塞的最高任务优先级,释放信号量后恢复该任务的优先级。优先级继承协议实现较为简单,但该协议无法检测死锁现象,必须由另外的机制实现死锁的预防或检测。

优先级上限协议是针对优先级继承协议无法避免死锁而提出的,该协议定义了临界资源的上限优先级,其值等于可能占用该资源的最高任务优先级。其根本思想是:若任务T抢占另一个占用临界资源S的任务T'而进入自己的临界区,则任务T的优先级必定高于临界资源S的上限优先级;如果该条件不能满足则T阻塞,T'继承T的优先级继续运行。优先级上限协议能有效的避免死锁,但引入了上限阻塞,是一种次优协议。

优先级上限协议实质是死锁预防策略和优先级继承协议的融合,以上限优先级的形式摒弃死锁条件中的“请求和保持”。假如任务T依次需要访问 S_1 、 S_2 和 S_3 资源,则在 S_2 或 S_3 被其它任务占用期间,T由于上限阻塞而无法访问尽管空闲的资源 S_1 ,也就是T必须等待所需的全部资源空闲时才能进入临界区。上限优先级协议虽然解决了死锁问题,但造成了资源的严重浪费,并延迟了进程的运行。此外,上述两种协议在纯优先级驱动的系统(如UC/OS-II)中很难实现,且临界资源的个数受优先级数目的限制^[2]。

4 阻塞调度算法的思想及其实现

阻塞调度算法针对优先级反转的第3个条件,通过修改任务调度算法,加入阻塞调度策略,直接调度阻塞最高优先级任务的从而避免优先级反转现象,不需调整任务优先级;而在申请临界资源时检测和解除死锁。

为了叙述的方便,假定可以用任务的优先级来标识任务,任务 T_i 的优先级为 i ,系统任务个数为 $N+1$,并且所讨论的阻塞都是因共享临界资源引起的。阻塞调度算法需要设置一个任务阻塞表Block[0..N],用于记录所有任务因访问临界资源而被阻塞的情况,初始值为null,若 T_i 被 T_j 阻塞则Block[i]=j。

4.1 wait和signal操作

wait(S)和signal(S)实现互斥信号量的P、V操作。信号量采用记录型结构,其成员lock标识信号量是否被锁定,初始值为false,prio保存锁定该信号量的任务优先级,初始值为null;L链表链接被阻塞的各个任务,初始值为null;其结构如下:

```
Type mutex = record
    lock boolean           ;信号量被锁定的情况
    prio integer          ;占用信号量的任务优先级
    L list of tasks       ;被阻塞的任务链表
end
```

在wait(S)操作中,若S没有上锁,则锁定信号量S并记录任务优先级,若S已经被锁定则在阻塞表Block中记录被阻塞的情况,但不修改该任务在系统就绪表中的就绪状态,仍参与任务调度。即不是通过系统就绪表来标识任务被互斥资源阻塞

的情况而是利用新增的阻塞表来标识任务阻塞状况。产生死锁的必要条件之一是“环路等待”,表现在阻塞表中为任务直接或间接被自己阻塞,形成阻塞环,因此可以检测死锁并及时处理,而不需要另外的死锁检测机制。wait操作算法如下,其中prio_cur保存着当前任务的优先级。

```
procedure wait (S)
    var S mutex ;
    var prio_cur integer = priority of task waiting S ;
    var prio_temp integer = prio_cur ;
    begin
        if ( Slock )then
            Slock = true ;  ;锁定信号量
            Sprio = prio_cur ; ;记录锁定该信号量的任务优先级
        else
            Block[prio_cur] = Sprio ; ;任务被 Sprio 阻塞
            while (Block[prio_temp] != null ) ;判定是否存在死锁环
                if (Block[prio_temp] = prio_cur )then
                    solve deadlock ; ;存在死锁环,解除死锁
                else
                    prio_temp = Block[prio_temp] ;
            endwhile
            add current task to SL ; ;加入 SL 阻塞链
            schedule () ; ;重新调度
        endif
    end
```

当任务释放信号量时执行signal(S)操作,若没有任务被阻塞则恢复信号量的初始状态,若有任务被阻塞,则直接或间接被阻塞的最高优先级任务就绪,并更新阻塞表.prio_prehighrddy中保存着系统最高就绪任务优先级,prio_highrddy保存着阻塞调度之后的最高就绪任务优先级(即执行signal操作的当前任务优先级)。若两者相等则最高优先级任务没有被阻塞,SL队列中最高优先级任务就绪,若两者不等则系统最高优先级任务被信号量S直接或间接堵塞,SL队列中阻塞最高优先级的任务就绪,尽管该任务可能不是SL队列中最高优先级任务。signal操作如下,其中prio_highblock中保存着SL队列中最高优先级任务或SL队列中阻塞系统最高优先级任务的阻塞

```
procedure signal (S)
    var S mutex ;
    var prio_cur integer = priority of current process signaling S ;
    var prio_prehighrddy integer = highest priority of ready tasks ;
    var prio_highblock integer ; ;即将就绪的任务优先级
    begin
        if SL = null then
            Slock = false ; ;释放信号量
            Sprio = null ; ;没有任务占用该信号量
        else
            if (Block[prio_prehighrddy] != prio_cur )then ; ;存在间接
                prio_highblock = prio_prehighrddy
                while (Block[prio_highblock] != prio_cur )
                    prio_highblock = Block[prio_highblock]
            else ; ;不存在间接阻塞
                prio_highblock = highest priority of taks in SL ;
```

表 1 执行时间对比表

测试内容	任务调度时间对比		没有反转的 P、V 操作时间		发生反转的 P、V 操作时间	
	常规调度	阻塞调度	上限协议	阻塞调度	上限协议	阻塞调度
时间/us	1.344 37	1.344 85	3.12×10^{-3}	2.26×10^{-3}	9.187×10^{-2}	8.92×10^{-3}

```

endif
Block[prio_highblock]=null ;任务就绪
remove ready task from SL ;
update (Block[0..N]) ;更新阻塞表,被 prio_cur

```

阻塞的

; 任务改为被 prio_high-

block 阻塞

endif

end

4.2 任务调度算法

全局变量 prio_highrdy 保存着系统就绪任务队列中最高任务优先级同时也保存在 prio_prehighrdy 全局变量中,以便在 signal 操作中判定最高就绪优先级任务是否被阻塞。接着调度算法判断该任务是否被互斥信号量阻塞,若被阻塞则直接调度阻塞该任务的任务,从而避免占用临界资源的低优先级任务被中间优先级任务抢占而导致优先级反转。整个算法的实现就是在正常任务调度中添加给出注释的两条语句。

```

procedure schedule ()
var prio_highrdy integer ;
begin
prio_highrdy = highest priority of ready tasks ;
prio_prehighrdy = prio_highrdy ;保存系统最高就绪优先级
while (Block[prio_highrdy] != null) ;如果被阻塞则直接调度
prio_highrdy = Block[prio_highrdy] ;阻塞该任务的任务
switch tasks ;
end

```

5 阻塞调度算法性能分析

阻塞调度算法主要是针对优先级上限协议中存在上限阻塞而严重影响系统资源利用的问题提出的,两者的性能对比分析如下。

5.1 资源利用率分析

优先级上限协议和阻塞调度算法的根本差别在于解决死锁策略的不同,优先级上限协议采取摒弃“请求和保持”条件的方法。但存在上限阻塞问题,不仅造成了资源的严重浪费,而且

可能延迟高优先级任务的执行时间^[4]。阻塞调度算法采用的是死锁检测和解除策略,没有引入新的阻塞,大幅度提高了资源利用率,并且能及时检测的死锁而不需其它检测机制。

5.2 执行时间分析

在通过 RTCA DO178B 标准认证的源码开放实时嵌入式系统 UC/OS-II 内核中测试优先级上限协议和阻塞调度算法的运行时间,测试平台为 WinXP P4 2.6 GHz 256 MB RAM,测试结果如表 1 所示。测试数据是通过多次重复调度计算而得到的,阻塞调度算法引入了阻塞调度,其调度时间稍有增加(约 5×10^{-4} us)。在没有发生优先级反转的情况下,两者的 P、V 操作时间相差不大,若发生优先级反转则两者的执行时间接近相差一个数量级,主要是因为上限协议需要反复修改任务就绪表和调整任务优先级造成的。

6 小结

阻塞调度算法依靠系统任务调度算法避免优先级反转,同时在申请临界资源过程中检测并解除可能发生的死锁,发现死锁时可采用剥夺资源或撤消进程的方法,具体根据系统需要确定。由于该算法采用死锁的检测和解除策略,其资源利用率高于采用死锁预防策略的优先级上限协议,并且不需反复调整任务优先级,其适用范围更为广泛,实现更简单。

(收稿日期 2007 年 3 月)

参考文献:

- [1] Sha L, Rajkumar R, Lehoczky J P. Priority inheritance protocols an approach to real-time synchronization[J]. IEEE Transactions on Computers, 1990, 39(9): 1175-1185.
- [2] Labrosse J. U/OS-II the real-time kernel[M]. 2nd ed. Beijing: Beijing University of Aeronautics & Astronautics Press, 2003.
- [3] 王永吉, 陈秋萍. 实时单调速率及其扩展算法的可调度性判定[J]. 软件学报, 2004, 15(6): 799-814.
- [4] 何东之, 李伟, 张向文. 一种新型实时调度算法研究[J]. 小型微型计算机系统, 2005, 26(11): 1965-1970.
- [5] 汤子瀛, 哲风屏, 汤小丹. 计算机操作系统[M]. 西安: 西安电子科技大学出版社, 2001.

6 结束语

本文对叙事性技术在战略态势表现中的应用进行了探讨,详细阐述了其中的关键技术: 基于词共现图的文本处理以及基于互增强的素材检索策略,实验结果证明该方法的有效性。

(收稿日期 2007 年 4 月)

参考文献:

- [1] Jones W M. On the adapting of political-military games for various purposes[R]. Rand Note, March 1986.
- [2] 司光亚, 胡晓峰. 一个面向战略决策的综合集成研讨环境[J]. 军事系统工程, 1999, 1: 2-7.
- [3] Dhiraj J, Wang J Z, Li J. The story picturing engine-a system for automatic text illustration [J]. ACM Transactions on Multimedia Computing, Communications and Applications, 2006.

(上接 17 页)

1000	87.17	76.31	56.12	62.18	58.84	66.82	83.31	37.13	35.83	53.65	34.89
87.17	1000	72.25	53.36	58.47	53.43	64.53	88.81	31.35	33.36	58.34	32.14
76.31	72.25	1000	51.89	59.63	51.53	64.74	61.48	36.03	31.29	49.7	38.24
56.12	53.36	51.89	1000	65.76	55.97	66.52	65.82	48.33	47.66	65.83	54.79
62.18	58.47	59.63	65.76	1000	78.82	65.27	84.81	57.86	57.2	78.98	68.75
58.84	53.43	51.53	55.97	78.82	1000	58.34	72.76	58.42	49.56	68.72	49.32
66.82	64.53	64.74	66.52	65.27	58.34	1000	88.2	39.4	44.28	64.89	49.28
83.31	88.81	61.48	65.82	84.81	72.76	88.2	1000	55.09	54.31	73.55	56.52
37.13	31.35	36.03	48.33	57.86	58.42	39.4	55.09	1000	58.25	71.39	71.79
35.83	33.36	31.29	47.66	57.2	49.56	44.28	54.31	58.25	1000	68.58	67.54
53.65	58.34	49.7	65.83	78.98	68.72	64.89	73.55	71.39	68.58	1000	74.58
34.89	32.14	38.24	54.79	68.75	49.32	49.28	56.52	71.79	67.54	74.58	1000

图 4 候选集视频相似度矩阵



图 5 候选集视频排序结果

(H1 表示最高重要性, L1 表示最低重要性)

一种多目标微粒群算法及其收敛性分析

王俊年^{1,2}, 刘建勋², 陈湘州³

WANG Jun-nian^{1,2}, LIU Jian-xun², CHEN Xiang-zhou³

1. 湖南科技大学 信息与电气工程学院, 湖南 湘潭 411201

2. 湖南科技大学 知识网格实验室, 湖南 湘潭 411201

3. 湖南科技大学 商学院, 湖南 湘潭 411201

1. School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, Hunan 411201, China

2. Knowledge Grid Laboratory, Hunan University of Science and Technology, Xiangtan, Hunan 411201, China

3. Economics School, Hunan University of Science and Technology, Xiangtan, Hunan 411201, China

E-mail: wangjunnian@sina.com

WANG Jun-nian, LIU Jian-xun, CHEN Xiang-zhou. Multi-objective particle swarm optimization and its convergence analysis. *Computer Engineering and Applications*, 2007, 43(22): 53-55.

Abstract: A stochastic multi-objective particle swarm optimization is proposed based on analyzing the multi-objective optimization problem. In this algorithm, the varieties of swarm are increased and obtained homogeneous Pareto front by selecting pbest and gbest of particle updating formula randomly from Pareto set repository. Then the convergence of the algorithm is analyzed by using finite Markov chain theory and it is proved that the SMOPSO algorithm converged to the minimizer by probability of 1.

Key words: multi-objective optimization; particle swarm optimization; Pareto dominance; Markov chain; convergence

摘要: 在分析多目标优化问题的基础上, 提出一种随机多目标微粒群算法, 该算法采用在已经获得的 Pareto 解集中随机选取的两个 Pareto 解作为微粒更新公式中的 pbest 和 gbest 微粒, 从而使微粒群的多样性增加, 获得均匀分布的 Pareto 前沿。之后利用有限齐次马尔科夫理论给出了 SMOPSO 算法的收敛性进行了分析, 证明 SMOPSO 算法以概率 1 收敛于极小元。最后通过对两个常用多目标函数的仿真实验, 验证了算法的有效性。

关键词: 多目标优化; 微粒群算法; Pareto 占优; 马尔科夫链; 收敛性

文章编号: 1002-8331(2007)22-0053-03 文献标识码: A 中图分类号: TP183

1 引言

多目标优化 (Multi-objective Optimization, MO) 问题是工程实践中广泛存在的一类复杂优化问题。经典的优化方法在多目标凸优化方面获得了广泛地应用, 如权重系数法等, 此类方法的关键在于确定权重系数, 但需要大量的先验知识或对初始种群中个体的分布作严格限制。基于群体及 Pareto 概念的多目标进化算法多目标进化算法 (Multi-objective Optimization Evolutionary Algorithms, MOEA) 能够很大程度上克服权重系数法的许多缺点, 如群体多样性差、先验知识不足及获得较少 Pareto 解等, 这类方法具有较高的并行性, 能有效地搜索 Pareto 解, 已经成为目前解决多目标优化问题的主要工具^[1]。微粒群算法从本质上讲也具有与进化算法同样的内在并行计算机制, 而且与进化算法相比, 微粒群算法具有算法概念简单、搜索效率高等特点, 因此, 将微粒群算法的应用领域扩展到多目标优化问题是很自然的事情。自 2002 年以来, 陆续出现了不同版本的多目标优化微粒群算法 (Multi-objective Optimizer PSO, MOPSO)^[2-10], 但与不断蓬勃发展的多目标进化算法 (MOEA) 比较, MOPSO 算

法无论在算法设计、理论分析和应用方面的发展都比较缓慢, 究其原因, 主要是因为 PSO 算法与 EA 算法不同的一些操作特征造成的, 在 PSO 算法运行过程中, 没有选择或变异操作, 因而微粒群的规模是固定的, 个体微粒不会被舍弃或者替换, 被替换的是迄今为止个体经历过的最优位置 (pbest) 和群体经历过的最优位置 (gbest), 群体中所有个体的调整均依赖于 pbest 和 gbest 的值。但是将 PSO 算法推广到求解 MO 问题时, 个体最优位置 pbest 和群体最优位置 gbest 分别变成了一组非受控制的解 (non-dominated solutions), 而不再有单独的 pbest 值和 gbest 值, 这时, 如何选择合适的 pbest 微粒和 gbest 微粒来引导群体内微粒的“飞行”就显得非常重要了, 目前开发的一些 MOPSO 算法主要的区别在于 pbest 微粒和 gbest 微粒的选择方法不同^[11], 且均没有进行必要的理论分析。进行必要的理论研究, 特别是对算法的收敛性、收敛速度等进行分析, 能够为设计高效的算法提供理论保障。

本文提出一种随机多目标微粒群算法 (Stochastic Multi-Objective PSO, SMOPSO), 并利用马尔科夫链理论对该算法的

基金项目: 国家自然科学基金 (the National Natural Science Foundation of China under Grant No.60474029)。

作者简介: 王俊年 (1968-), 男, 博士, 副教授, 研究方向: 智能优化算法与智能控制。

收敛性(收敛于一个 Pareto 最优解集)进行了分析,对两个典型的多目标函数优化问题的仿真实验表明,本文的算法能够获得分布比较均匀 Pareto 前沿。

2 预备知识

定义 1 (多目标优化问题 MOP) 设 R^n 为 n 维欧氏空间 D , $R^n \times D \times R^m$ 为目标函数 $g(x)$, $h_j(x)$ 均为 D 上的约束函数,含约束多目标极小化问题可描述为

$$\min_{x \in D} f(x) = (f_1(x), f_2(x), \dots, f_n(x))$$
$$g(x) \leq 0, h_j(x) = 0, j = 1, 2, \dots, l, j = 1, 2, \dots, m$$

定义 2 (Pareto 占优) 设 $x = (x_1, x_2, \dots, x_n)$, $y = (y_1, y_2, \dots, y_n)$, 若 $\forall i \in \{1, 2, \dots, n\}$, $x_i \leq y_i$, 且至少存在一个 i , 使得 $x_i < y_i$, 则称 x 比 y 占优。

定义 3 (Pareto 最优解) 对于 $\forall x \in D$ 和 $J = \{1, 2, \dots, m\}$, 如果存在 $x^* \in D$, 使得

$$\forall j \in J, f_j(x^*) \leq f_j(x)$$

且至少存在一个 $j \in J$, 使得

$$f_j(x^*) < f_j(x)$$

则称 $x^* \in D$ 是 MO 问题的一个 Pareto 最优解, 亦称 Pareto 解。

定义 4 (Pareto 最优解集) 多目标优化问题 $f(x)$ 的所有 Pareto 解的集合, 称为 $f(x)$ 的 Pareto 解集, 记作 P^* 。

定义 5 (Pareto 前沿) 对于给定的多目标优化问题 $f(x)$ 及其 Pareto 解集 P^* , Pareto 前沿 (Pareto front) 定义为

$$PF = \{f(x) = (f_1(x), f_2(x), \dots, f_n(x)) | x \in P^*\}$$

定义 6 (极小元) 设 S 是严格偏序集 $(S, <)$ 的子集 a , 叫做 a 的极小元, 是指不存在 $x \in S$, 使 $x < a$ 。定义集合 $M(S, <) = \{a \in S | a \text{ 是 } S \text{ 的极小元}\}$, 称 $M(S, <)$ 为关于 $(S, <)$ 的极小元集; 称极小元集 $M(S, <)$ 是完全的, 如果对于每一个 $x \in S$, 都至少有一个 $x' \in M(S, <)$ 满足 $x' < x$ 。

定义 7 (有限齐次马尔科夫链) 如果 S 是一个有限集, 并且 $\{X_t, t \in N_0\}$ 是一个从 S 中取值的随机序列, 对于所有的 $t \in N_0$ 和所有的 $(i, j) \in S \times S$ 有以下性质:

$P\{X_{t+1}=j | X_t=i, \dots, X_0=i_0\} = P\{X_{t+1}=j | X_t=i\} = P_{ij}$, 则称序列 $\{X_t, t \in N_0\}$ 是一个在状态空间 S 上的齐次有限马尔科夫链。

定义 8 (不可约随机矩阵) 如果一个非负矩阵的每一行元素之和等于 1, 则称它是随机的。称随机矩阵 P 是不可约的, 如果有 $\forall i, j \in S: \exists k \in N, P_{ij}^{(k)} > 0$ 。

定理 1 [12,13] 一个具有有限状态空间与不可约转移矩阵的齐次马尔科夫链, 以概率 1 无限此地访问每一状态而与它的初始分布无关。

3 随机多目标 PSO

与目前已经提出的一些 MOPSO 算法不同, 本文的 SMOPSO 算法只有单一的微粒群, 且只设置一个存储区 r_{best} , 用于存储迄今为止搜索到的 Pareto 解。微粒群进化时在 Pareto 解存储区 r_{best} 中分别随机选择一个 Pareto 解作为 $pbest$ 和 $gbest$, 第 i 个微粒在第 j 维上的速度和位置更新公式如下

$$v_{ij}(t+1) = \omega v_{ij}(t) + c_1 r_1 [y_{ij}(t) - x_{ij}(t)] + c_2 r_2 [\hat{y}_j(t) - x_{ij}(t)] \quad (1)$$
$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad (2)$$

式中 y_{ij}, \hat{y}_j 分别是当前已获得的 Pareto 解集 r_{best} 中随机选取的一个 Pareto 解, 其他参数 $\omega, c_1, r_1, c_2, r_2$ 与标准微粒群算法中微粒的更新公式中相同。

SMOPSO 算法在处理 X_i 的边界问题时与标准微粒群算法

处理 X_i 边界的方法不同, 在 SMOPSO 中, 当 x_{ij} 不满足约束条件时, 采用对 x_{ij} 进行重新初始化的方法来处理, 有利于增加搜索群体的多样性, 防止早收敛发生。SMOPSO 算法流程如下:

步骤 1 初始化

初始化微粒的初始位置和初始速度, 以及式(1)中的参数 $\omega, c_1, r_1, c_2, r_2$ 。

(1) 随机初始化微粒群中个体微粒的位置和速度, 检验各微粒是否满足约束条件, 若不满足, 则对其重新初始化;

(2) 计算每一个个体微粒的多目标适应值并检查其 Pareto 占优情况, 将非劣解微粒存储于 Pareto 解集存储区 r_{best} 中。初始化的 r_{best} 中至少需要有两个微粒, 若不足则随机产生。

步骤 2 微粒群适应值评价

计算微粒群中各微粒个体的每一个目标适应值, 并以矢量形式存储。

步骤 3 微粒更新

(1) 用公式(1)、(2)更新微粒群中个体微粒的速度和位置, 其中 y_{ij}, \hat{y}_j 从当前 Pareto 解集中任意选取一个。

(2) 如果更新后的微粒不满足约束条件, 则将该微粒删除后重新随机产生一个。

步骤 4 更新 Pareto 解存储区

(1) 检查微粒群中每一个微粒的 Pareto 占优情况, 若某个占优微粒和 Pareto 解集 r_{best} 中的微粒相比较是非劣解的话, 则将该微粒存入 r_{best} 作为当前获得的一个 Pareto 解。

(2) 如果 Pareto 解集 r_{best} 中某个微粒受控于新存入的微粒, 则从 Pareto 解集 r_{best} 中删除该微粒。

步骤 5 循环

循环执行(2)到(4)直到符合算法停止条件。HPSO 算法中停止条件为迭代次数等于预设最大迭代次数。

4 SMOPSO 算法的收敛性分析

利用有限马尔科夫链进行 MOPSO 收敛性分析的基本思想是 (1) 证明算法在执行过程中, 其微粒群序列是一个具有有限状态空间和不可约转移矩阵的齐次马尔科夫链。这样, 定理 1 保证微粒群可以以概率 1 遍历每一个状态, 当然也包括了包含 Pareto 最优解的状态; (2) 要保证一旦获得了 Pareto 最优解, 算法应该具有某种机制把最优解保存下来; (3) 还要证明作为结果输出的集合中所有非最优解能以概率 1 被淘汰。

定义 9 (MOPSO 收敛性) 令 A_t 是 MOPSO 算法在第 t 代的微粒群, $F_t = f(A_t)$ 是其相应的适应值矢量(目标空间)集合, F^* 是所求问题目标空间集合的极小元集, 则如果 $\lim_{t \rightarrow \infty} P(F_t) = 0$ 以概率 1 成立, 则称 MOPSO 算法以概率 1 收敛于极小元。其中, $\alpha(A) = |A| - |A \cap B|$ 。

结论 1 SMOPSO 算法的微粒群序列 $(A_t), t \in N_0$ 是具有正转移矩阵的齐次有限马尔科夫链。

证明 若算法求解的是一个连续问题, 但在确定的精度意义下, 仍可以把状态空间 S 看作有限集[12]。由于 SMOPSO 算法的微粒的迭代公式(1)中 y_{ij}, \hat{y}_j 从 Pareto 解集 r_{best} 中随机选取, 每一个微粒的更新是独立进行的, 新微粒群仅与上一代微粒群及 y_{ij}, \hat{y}_j 有关, 而与上一代微粒群之前的群体无关, 即群体无后效性, 并且各代微粒群之间的转换概率与时间的起点无关, 即

$$P\{X_{t+1}=j | X_t=i, X_{t-1}=i_{t-1}, \dots, X_0=i_0\} = P\{X_{t+1}=j | X_t=i\} = P_{ij}$$

根据定义 7, SMOPSO 算法的微粒群序列 $(A_t), t \in N_0$ 的转移矩

阵 P 与时间 t 无关的齐次有限马尔科夫链。

接下来需要证明转移矩阵是正矩阵。SMOPSO 算法中微粒群序列的转移矩阵主要涉及微粒群更新式(1)和式(2),因速度更新公式中存在的 $r_1(t)$, $r_2(t)$ 项以及 $y_{i1}(t)$, $\hat{y}_{j1}(t)$ 在 Pareto 解集中的随机选择性,使得

$$P(v_{ij}(t) > 0) > 0$$

因此,通过式(1)和式(2)这两步的迭代执行,一个个体微粒在下一代存在变化成任意一个个体(即解空间中任意位置)的可能性,因而它的转移矩阵是正的。

综上所述可以得出,SMOPSO 算法中的微粒群序列 $(A_t)_t=0$ 是具有正转移矩阵的齐次有限马尔科夫链。

结论 2 如果 MOPSO 算法的微粒群序列 $(A_t)_t=0$ 是具有正转移矩阵的齐次有限马尔科夫链,那么该算法以概率 1 收敛于极小元集。

证明 设 $f(A_t)$ 是 A_t 的适应值向量集(像集), $\hat{M} = M(\cdot, \cdot)$ 是问题的极小元集,且是完全的(即 Pareto 最优解集)。

在 SMOPSO 算法的执行过程中,“排序符号” $<$ 的应用其实就是步骤 4 中非劣解(Nondominance)的求解过程,找到的非劣解将进入 Pareto 最优解集 r_{best} 。算法迭代的最终结果是 $f(r_{best})$ 。

这一点将从以下两个方面证明。

首先,要证明一个最优解一旦被找到进入 Pareto 解集存储区 R_{best} 就永远不会被淘汰。SMOPSO 算法的步骤 4(2)说明,当主微粒群中新找到的一个非劣解与 Pareto 最优解存储器 R_{best} 中的某个微粒群 a 相比较如果占优(dominated)的话,则微粒 a 从 r_{best} 中被删除,亦即一个微粒 a $\in r_{best}$ 被删除,当且仅当存在一个非劣解 b $\in A_t$ 使得 $f(b) < f(a)$ 。因此,根据 $\hat{M} = M(\cdot, \cdot)$ 的定义, \hat{M} 中的微粒一旦进入 r_{best} 就永远不会被删除。

其次,要证明进入 R_{best} 的非最优个体在有限代内必将被删除。令微粒 a $\in r_{best}$ 且 $f(a) \notin \hat{M}$ 。根据定义 7,一定存在一个微粒 b $\in A_t$ 满足 $f(b) < f(a)$ 。根据 r_{best} 的更新准则,只要这个更好的微粒能经过一定的迭代次数 t 在 $(A_t)_t=0$ 中出现,它一定会进入 r_{best} 取代 a。这一过程不断发生,直到 r_{best} 中没有非最优微粒。由于 $(A_t)_t=0$ 是具有正转移矩阵的齐次有限马尔科夫链,定理 1 保证了解空间中的所有位置都会以概率 1 被无限次访问。因此, r_{best} 中的非最优微粒经过有限次迭代以后会以概率 1 淘汰。

综上所述,所有进入 r_{best} 的最优解都不会被淘汰,并且所有 r_{best} 中的非最优解都会以概率 1 在有限步内被淘汰。因此当 $T \rightarrow \infty$ 时 $f(r_{best}) \subset \hat{M}$, 因而有 $\lim_{t \rightarrow \infty} P(F_t) = 0$ 以概率 1 成立。根据定义 9, SMOPSO 算法以概率 1 收敛于极小元。

5 算法测试

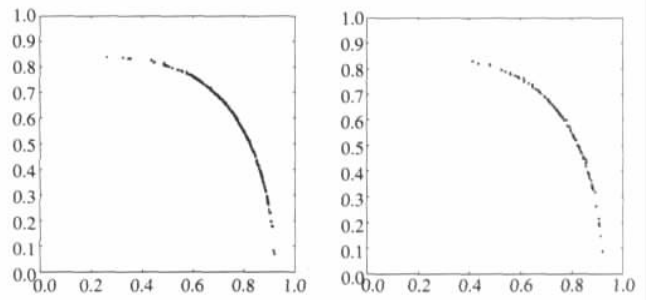
选择三个典型的测试函数对提出 HPSO 算法与文献[6]提出的动态邻域 PSO 算法(DNPSO)进行对比测试。公式(1)中的参数分别选择如下: $\omega = 0.5 + \text{rand}$, $c_1 = c_2 = 1.5$, 微粒群的规模 $\text{pop-size} = 50$ 。

测试函数 1 首先选择的测试函数为

$$\min F_1 = (f_1(x, y), f_2(x, y))$$

其中 $f_1(x, y) = \sqrt{x^2 + y^2}$, $f_2(x, y) = \sqrt[4]{(x-0.5)^2 + (y-0.5)^2}$, $x, y \in [0, 10]$ 。

该问题是具有连续 Pareto 前端的无约束多目标优化问题,分别运行 200 代以后, SMOPSO 算法找到了 188 个 Pareto 解,如图 1(a)所示;文献[6]的 DNPSO 算法找到了 78 个 Pareto 解,如图 1(b)所示。



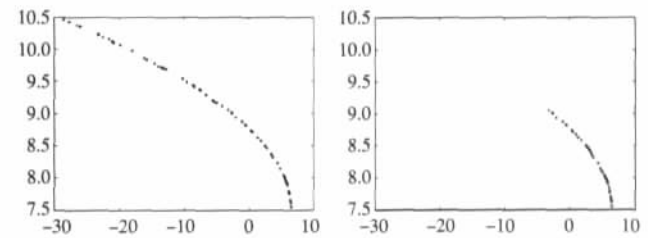
(a) SMOPSO 产生的 F_1 Pareto 前端 (b) DNPSO 产生的 F_1 Pareto 前端
图 1 两种算法产生的 F_1 函数 Pareto 前端比较

测试函数 2 选择的第 2 个测试函数为^[14,15]

$$\max F_2 = (f_1(x, y), f_2(x, y))$$

其中 $f_1(x, y) = -x^2 + y$, $f_2(x, y) = \frac{1}{2}x + y + 1$, 约束条件为 $0 \leq \frac{1}{6}x + y \leq \frac{13}{2}$, $0 \leq \frac{1}{2}x + y - \frac{15}{2} \leq 0$, $5x + y - 30 \leq 0$, $x, y \geq 0$ 。

这是一个含有约束条件的多目标优化问题,与测试函数 1 的条件相同,两种算法对测试函数 2 运行 200 代以后得到的 Pareto 前端如图 2(a)和图 2(b)所示。



(a) SMOPSO 产生的 F_2 Pareto 前端 (b) DNPSO 产生的 F_2 Pareto 前端
图 2 两种算法产生的 F_2 函数 Pareto 前端比较

由 SMOPSO 算法和 DNPSO 算法对两个典型函数 F_1 、 F_2 的测试结果(图 1 及图 2)看到, SMOPSO 不仅得到了较多的 Pareto 解,而且其得到的 Pareto 前端分布也比 DNPSO 得到的 Pareto 前端分布广。

6 结论

本文提出一种随机多目标微粒群算法,简化了多目标微粒群算法中微粒更新公式中微粒个体目前为止经历的最优位置 $y_{i1}(t)$ 和群体目前为止经历的最优位置 $\hat{y}_{j1}(t)$, 而且 $y_{i1}(t)$, $\hat{y}_{j1}(t)$ 的随机选择性也增加了算法微粒的多样性,使获得的 Pareto 前端分布均匀。应用有限齐次马尔科夫理论给出了 SMOPSO 算法的收敛性证明,是对多目标 PSO 算法理论分析的有益探索,通过 SMOPSO 算法的收敛性分析测试实验也发现,在 MOPSO 算法的 $y_{i1}(t)$, $\hat{y}_{j1}(t)$ 选取策略中需要保持算法搜索的遍历性(以满足马尔科夫性)和算法效率之间的矛盾平衡是 MOPSO 算法设计中需要进一步考虑的问题。(收稿日期:2007 年 3 月)

参考文献:

- [1] Zitzler E, Deb K, Thiele L. Comparison of multi-objective evolution algorithms: empirical results[J]. Evolutionary Computation, 2000, 8(2): 173-195.
- [2] Parsopoulos K E, Vrahatis M N. Particle swarm optimization method in multiobjective problem[C]// Proceedings ACM Symposium on Applied Computing, Leganes, Spain, 2002: 603-607.

(下转 92 页)

基于统计特征的图像修复

陈芳¹,叶正麟¹,王继红²

CHEN Fang¹,YE Zheng-lin¹,WANG Ji-hong²

1.西北工业大学 理学院,西安 710072

2.西北工业大学 软件与微电子学院,西安 710072

1.School of Science, Northwestern Polytechnical University, Xi'an 710072, China

2.Software and Microelectronics College, Northwestern Polytechnical University, Xi'an 710072, China

E-mail: wjh_cf@yahoo.com.cn

CHEN Fang, YE Zheng-lin, WANG Ji-hong. Image inpainting based on statistical features. Computer Engineering and Applications, 2007, 43 (22): 56-58.

Abstract: An image inpainting method based on the existing exemplar-based image inpainting idea is proposed. Improves the method by the filling order and matching criterion. The priority is added in gradient information and variance information. It makes sure that the filling order can propagate texture and structure information correctly. The matching criterion is added in the color eigenfunction and statistical eigenfunction. Proposes a two-pass matching criterion to get the optimum matching patch. A number of examples demonstrate the effectiveness and the robustness of our algorithm.

Key words: image inpainting; texture synthesis; color eigenfunction; statistical eigenfunction

摘要: 在基于样图图像修复思想的基础上,从填充顺序和匹配准则两个方面进行改进,提出了一种图像修复方法。优先级函数加入了面片梯度信息和方差信息,使得填充顺序能够正确地修复结构和纹理信息。匹配准则采用自定义的颜色特征函数和统计特征函数对相似片进行两重匹配,取最优匹配块填充。一系列的修复实例证明该方法的有效性和鲁棒性。

关键词: 图像修复; 纹理合成; 颜色特征函数; 统计特征函数

文章编号: 1002-8331(2007)22-0056-03 文献标识码: A 中图分类号: TP391

1 引言

图像修复是利用图像中待修复区域周围已知像素的信息,来填充待修复区域,使得整幅图像达到视觉上的连续和完整。待修复区域指图像中的受损区域或多余物体,如斑点、文字、褶皱、障碍物等等。图像修复在图像处理、图像分析、电影工业、图像传输等中有着广泛地应用。目前图像修复中占主流的修复模型有(1)偏微分方程的修复模型^[1-4]。(2)纹理合成的修复模型^[5-10]。前者适用修复那些待修复区域较窄的情况,针对性太强,适用范围偏窄。相比之下,后者是近几年发展起来的,它适用范围较广,修复效果好。

基于样图的图像修复方法^[7]有效地结合偏微分方程和纹理合成修复的优点,这类方法一般分为3步:(1)计算优先级,确定当前需要填充的面片(简称当前面片)。优先级是指面片填充的顺序,优先级高的面片先填充,优先级低的后填充。(2)根据当前面片中已知信息,在图像的已知区域寻找相似块。(3)用相似块填充当前面片。

Criminisi^[7]修复效果比较理想,但还存在以下问题:(1)优

先级函数的定义不够理想, $P(\phi)=C(\phi)*D(\phi)$,当 $D(\phi)=0$ 时, $P(\phi)=0$,这时即使 $C(\phi)$ 很大,甚至整个面片中只有几个像素未知,面片也得不到及时的填充。(2)匹配准则不准确。面片中已知像素信息是其选择填充面片(即相似块)的约束条件。一个面片在局部源区域中往往能找到多个相似块,它们颜色相近但内容往往相差很远。这时如果按传统的修复方法,任意选取一块来填充,容易产生匹配误差,误差的传播和累计,会导致最终结果错误。所以单靠颜色信息来判断两个面片是否相似是远远不够的。

图像修复首先要保证的一点是结构连线在待修复区域内应该是连续的,所有的修复模型都建立在这个前提之下。人眼认识事物的初级阶段是物体的形状(即图像的结构信息),而视觉对一定范围内颜色(即图像的纹理信息)的变化反应是迟钝的。因此确定一个准确的优先级函数和描述能力强的匹配准则是一个好修复方法的关键。

基于以上的考虑,本文在充分利用图像统计特征的基础上,提出了一种改进的图像修复方法。包含像素梯度信息和方

基金项目:国家自然科学基金(the National Natural Science Foundation of China under Grant No.60672135);西北工业大学研究生创业种子基金项目(No.Z200656)。

作者简介:陈芳(1982-),女,硕士研究生,研究方向为图像处理;叶正麟(1943-),男,教授,博士生导师,主要研究方向:计算几何、计算机图形学、图像处理;王继红(1981-),男,硕士研究生,研究方向为集成电路可测试性设计、数字信号处理。

差信息的优先级函数,保证填充顺序能够正确地修复结构信息和纹理信息。自定义的匹配准则,利用颜色匹配函数和统计匹配函数对相似片两次筛选,取最优匹配块填充。

2 自定义优先级函数和匹配准则

设 Ω 为待修复区域,也称目标区域或未知区域 (Ω 中的像素值未知), $\partial\Omega$ 为待修复区域边界。I 为整幅图像,称为源区域,是相似块的来源。 p 代表以点 p 为中心的边缘正方形面片 $p \in \partial\Omega$, 设正方形的尺寸为 $w \times w$, 本文一般取 9×9 像素。 Φ 为以点 p 为中心的大的正方形区域,为局部源区域。搜索相似块的时候只在 Φ 中进行,而不在 I 中全局搜索。 Φ 的大小一般为 $2w \times 2w$ 。

2.1 面片优先级函数 $P(p)$ 的定义

$P(p)$ 的定义如下:

$$P(p) = \begin{cases} 0, & \text{若 } C(p) < 0.5 \\ C(p) + D(p), & \text{若 } C(p) \geq 0.5 \end{cases}$$

其中 $C(p), D(p)$ 的定义如下:

$$C(p) = \frac{N}{w \times w} \quad D(p) = \text{grad}(p) + \text{sstd}(p)$$

这里 N 为面片 p 内已知像素的个数, $C(p)$ 表示 p 中含已知像素的百分比。 $\text{grad}(p), \text{sstd}(p)$ 分别反映 p 内的平均梯度信息和方差信息, $\text{grad}(p)$ 越大,表明 p 内纹理的方向性越强,含有的结构信息越多; $\text{sstd}(p)$ 越大,表明 p 内颜色分布不均匀的程度越高。特别地,当 $\text{sstd}(p) = 0$ 时,说明 p 内颜色分布比较均匀,颜色十分接近,这也和视觉保持一致。 $\text{grad}(p), \text{sstd}(p)$ 的计算如下:

$$\text{grad}(p) = \frac{G_x + G_y}{255} \quad \text{sstd}(p) = \frac{\sigma(r) + \sigma(g) + \sigma(b)}{\mu(p)}$$

这里 $G_x + G_y$ 是 p 内平均梯度,本文中 σ^2 表示方差, μ 表示期望(即颜色均值),计算式如下:

$$G_x = \frac{1}{N} \sum_{i=1}^N \sqrt{\left(\frac{\partial r_i}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial b_i}{\partial x}\right)^2}$$

$$G_y = \frac{1}{N} \sum_{i=1}^N \sqrt{\left(\frac{\partial r_i}{\partial y}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2 + \left(\frac{\partial b_i}{\partial y}\right)^2}$$

$$\sigma(r) = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i - \mu(r))^2}$$

$$\mu(p) = \frac{1}{N} \sum_{i=1}^N (r_i(p) + g(p) + b_i(p))$$

当是灰度图像时, G_x, G_y 相应地是对灰度差分求平均值。这样定义 $P(p)$ 有三个好处: (1) 改善了图像修复的效果。当 $C(p) < 0.5$ 时,面片 p 内包含的已知像素不到一半,强行修复容易产生人工痕迹。(2) 提高了算法的速度。当 $C(p) < 0.5$ 时,不考虑修复 p ,这时就不计算 $D(p)$,为算法运行节约了大量的时间。(3) 最大颜色梯度不足以有效反映面片内所有已知像素的梯度信息,这里颜色梯度的平均值代替其最大值,就可以避免这个缺陷,同时还可以抑制噪声的影响。

2.2 匹配准则 Dist 的定义

由于颜色信息是一种重要的视觉信息属性,它很稳定;另一方面颜色信息又丢失了图像的空间关系,可以说图像的信息并没有完全利用。所以本文提出如下匹配准则 Dist ,引入颜色特征函数和统计特征函数,在考虑图像颜色信息的同时,充分利用整个面片的全局统计特征,使得像素间的相对位置关系得到考虑,从空间关系上弥补了传统方法的不足之处。

$\text{Dist}(p, q) = k \cdot \text{diff}(p, q) + \text{CZD}(p, q)$ (k 为权因子,本文取 $k=1$)

其中 diff 是颜色特征函数,计算如下:

$$\text{diff}(p, q) = \text{MAD}(p, q) + \text{mMPC}(p, q)$$

$$\text{MAD}(p, q) = \frac{1}{N \mu(p)} \sum_{i=1}^N (|r_i(p) - r_i(q)| + |g(p) - g(q)| + |b_i(p) - b_i(q)|)$$

$$\text{mMPC}(p, q) = 1 - \frac{1}{N} \sum_{i=1}^N t_i$$

$$t_i = \begin{cases} 1 & \text{若 } |r_i(p) - r_i(q)| + |g(p) - g(q)| + |b_i(p) - b_i(q)| \\ & (r_i(p) + g(p) + b_i(p)) \\ 0 & \text{其它} \end{cases}$$

即 MAD 是比较两个面片的平均颜色绝对差, mMPC 是统计 p 中不匹配的像素的比例。

CZD 是统计特征函数,计算如下:

$$\text{CZD}(p, q) = a + b + c$$

$$a(p, q) = \frac{|\sigma(p) - \sigma(q)|}{\sigma(p)} \quad b(p, q) = \frac{|\mu(p) - \mu(q)|}{\mu(p)}$$

$$c(p, q) = b(p_1, q_1) + b(p_2, q_2) + b(p_3, q_3) + b(p_4, q_4)$$

其中 a, b 分别用于比较两个面片 p, q 的方差差异和颜色均值差异,方差 $\sigma(p)$,期望 $\mu(p)$ 是归一化因子。 c 表示将面片 p, q 各自按左上、右上、左下、右下 4 等分后,4 个对应子块之间灰度均值的差异。

3 算法的流程图

搜索相似块的原则:设定误差阈值,用 diff 和 Dist 函数对候选面片进行两次筛选,取最优匹配块来填充。 $d(p) =$

$\frac{1}{N \mu(p)} \sum_{i=1}^N (r_i(p) + g(p) + b_i(p))$, 这里 ρ 是用户指定的, $\rho < 1$, 本文中取 $0.1 \sim 0.4$ 。本文规定当 $\text{diff}(p, q) \leq \rho \cdot d(p)$ 时,则认为面片 q 为当前面片 p 的相似块。对这些相似块计算其 $\text{Dist}(p, q)$, 最优匹配块 q^* 满足: $q^* = \arg \min_{q \in \Phi} \text{Dist}(p, q)$

算法的流程图如图 1。

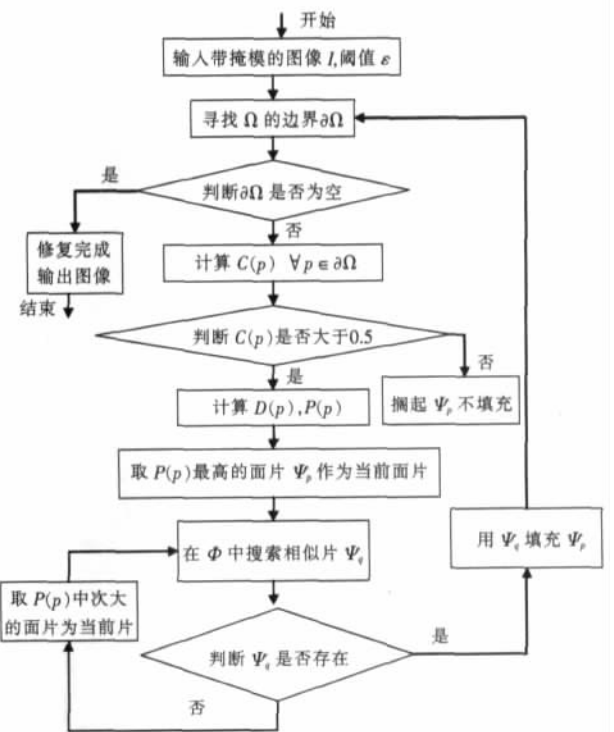


图 1 算法流程图

4 实验结果

算法以 MATLAB 为平台, 在 P III 256 MB 内存的微机实现。用户只需要手动标出待修复的区域, 剩下的将全部由算法自动完成。

图 2 和图 6 是去除图像中的划痕, 修复区域包含很强的结构信息, 图 3 中含多个小的斑点修复区域, 既有纹理修复又有结构修复, 图 4 和图 5 是去除图像中的多余物体。通过图 2、图 3 将 Criminisi 方法和本文方法的比较, 可以看出本文算法有效



图 2 算法改进前后实验结果的比较



图 3 算法改进前后实验结果的比较



(a)原图(396x325) (b)本文结果

图 4 移出图像中的多余物体



(a)原图(417x580) (b)本文结果

图 5 去掉图像中的文字



(a)原图(256x256) (b)待修复图像 (c)本文结果

图 6 去掉图像中的划痕

地避免了圆圈中的人工痕迹。

5 结束语

本文提出了一种图像修复的方法, 从以下几个方面提高了算法的有效性和鲁棒性: (1) 在优先级函数中, 加入了方差和平均梯度信息, 并规定待填充的面片中含已知像素的百分比不低于一半, 提高了结构信息和纹理信息的准确程度。 (2) 相似块的选择采用两重匹配, 自定义的匹配准则加入了颜色和方差、期望等信息, 极大地改善了图像修复的效果。

进一步的工作可以从以下几个方面考虑: (1) 算法每填充完一个边缘面片, 就更新整个待修复区域的边界 $\partial\Omega$, 计算所有边缘面片的优先级, 这个过程有待优化。 (2) 目前图像修复技术还是一种交互式的修复模型, 即先由用户标出待修复区域, 然后由算法来完成剩下的修复工作。全自动的修复方法等待开辟。(收稿日期 2007 年 3 月)

参考文献:

- [1] Bertalmio M, Sapiro G, Caselles V, et al. Image inpainting[C]//Proc ACM Conf Comp Graphics (SIGGRAPH), New Orleans, LA, July 2000: 417-424.
- [2] Ballester C, Bertalmio M, Caselles V, et al. Filling-in by joint interpolation of vector fields and grey levels[J]. IEEE Trans Image Processing, 2001, 10: 1200-1211.
- [3] Bertalmio M, Bertozzi A L, Sapiro G. Navier-Stokes fluid dynamics, and image and video inpainting[C]//Proc Conf Comp Vision Pattern Rec, Dec 2001: 355-362.
- [4] Bertalmio M, Vese L, Sapiro G, et al. Simultaneous structure and texture image inpainting[C]//Proc Conf Comp Vision Pattern Rec, Madison, WI, 2003.
- [5] Efros A, Leung T. Texture synthesis by nonparametric sampling[C]//Proc Int Conf Computer Vision, Kerkyra, Greece, Sept 1999: 1033-1038.
- [6] Liang L, Liu C, Xu Y, Q, et al. Real-time texture synthesis by patch-based sampling[C]//ACM Trans Graphics, 2001.
- [7] Antonio Criminisi, Patrick Pérez, Kentaro Toyama. Region filling and object removal by exemplar-based image inpainting[J]. IEEE Transactions on Image Processing, 2004, 13(9).
- [8] Li Bian-ru, Qi Yue, Shen Xu-kun. An image inpainting method[C]//IEEE Ninth International Conference on Computer Aided Design and Computer Graphics, 2005.
- [9] Drori I, Cohen-Or D, Yeshurun H. Fragment-based image completion[C]//ACM Trans Graphics (SIGGRAPH), San Diego, CA, 2003: 303-312.
- [10] Jia J, Tang C K. Image repairing: robust image synthesis by adaptive ND tensor voting[C]//Proc Conf Computer Vision and Pattern Recognition, Madison, 2003.

基于神经网络的交互式图像检索

王兵¹, 张欣², 王硕¹, 王苗¹

WANG Bing¹, ZHANG Xin², WANG Shuo¹, WANG Miao¹

1. 河北大学 数学与计算机学院, 河北 保定 071002

2. 河北大学 电子信息工程学院, 河北 保定 071002

1. College of Mathematics and Computer Science, Hebei University, Baoding, Hebei 071002, China

2. College of Electronics and Information Engineering, Hebei University, Baoding, Hebei 071002, China

E-mail: wangbing@cmc.hbu.cn

WANG Bing, ZHANG Xin, WANG Shuo, et al. Interactive method of image retrieval based on neural networks. Computer Engineering and Applications, 2007, 43 (22): 59-62.

Abstract: An interactive image retrieval based on neural networks is employed to measure the similarities between the query image and the images in a database. The neural networks presented in this paper are constructed dynamically according to the end-users' relevance feedbacks. These similarities and historical relevance feedbacks deposited in the neural network can be used in next retrievals to gain enhancement in the efficiency of the system retrieval. This approach may be embedded in many current image retrieval systems to improve the performance of the retrieval. Results in our experimentation demonstrate a good potential of the methodology.

Key words: neural networks; content-based image retrieval; relevance feedback; set partition

摘要: 提出了基于神经网络的交互式图像检索方法, 系统根据用户对检索结果的评价, 动态构造神经网络, 描述图像之间的相似性, 图像间的这种相似性以及本次检索结果可以作为以后检索的历史信息保存在神经网络中, 从而提高下一次检索的效率。实验表明, 该方法嵌入到典型的图像检索系统中, 改善了图像检索性能。

关键词: 神经网络; 基于内容的图像检索; 相关反馈; 集合的划分

文章编号: 1002-8331(2007)22-0059-04 文献标识码: A 中图分类号: TP391.41

1 引言

基于内容的图像检索 (CBIR) 已经成为支持有效的信息搜索和信息浏览的一项具有实用价值的关键技术。它在数据库、信息检索和计算机视觉等领域受到广泛地关注。基于内容的图像检索是指从图像中直接提取出对象的语义、特征 (如图像的颜色、纹理、形状等), 然后根据这些线索从大量存储在数据库中的图像中进行查找, 检索出具有相似特性的图像来^[1]。在过去的十几年里, 人们建立了许多有效的基于内容图像检索系统, 其中一些被成功地应用在诸如生物医学、商业、教育和 Web 搜索引擎等领域上^[2-7]。

然而, 目前大多数基于内容的图像检索系统在实际应用中还不十分令人满意, 究其主要原因是图像低层特征和图像所表达的高层概念 (图像语义) 之间的一致性导致系统性能难以提高。称这种不一致性为“语义鸿沟”^[8]。根据用户在检索过程的参与程度, 目前主要有两类技术用于减少语义鸿沟, 一是相关反馈技术, 二是基于统计分类的图像数据库预处理技术^[9]。根据用户的反馈, 系统更新相关信息, 如调整特征权值、重新构造查询向量, 修改相关的先验概率等, 从而在下一步检索时得到用

户更多的主观信息。统计分类方法是根据图像低层视觉特征将图像按语义含义进行聚类, 以便构造适应于每一类别的语义检索方法。

典型的 CBIR 系统根据检索结果图像与查询图像的相似程度对检索结果进行排序, 输出显示, 而忽略了图像数据库中图像间所固有的相关性和未能利用检索的历史信息。本文针对这些问题进行研究, 提出了基于神经网络的交互式图像检索方法, 系统获得用户反馈信息, 以此动态构造神经网络, 描述图像之间的相似性, 图像间的这种相似性以及本次检索结果可以作为以后检索的历史信息保存在神经网络中, 从而提高下一次检索的效率, 改善图像检索性能。

2 基于神经网络的图像检索

2.1 系统概述

图 1 描述基于神经网络的图像检索系统的工作过程。系统先提取图像的特征, 并用特征向量描述一幅图像的内容, 这一过程称为特征提取。图像的特征可以事先提取出来, 存储在图像特征数据库里。然后, 在图像库预处理阶段对图像库中图像

基金项目: 河北省科技公关计划 (the Key Technologies R&D Program of Hebei Province, China under Grant No.072135139); 河北省教育厅项目 (No.Z2006414); 河北省省级精品课建设项目。

作者简介: 王兵 (1966-), 女, 副教授, 主要研究方向: 图像和文字处理。

进行预分类。利用图像特征和图像相似性测量方法测量图像间的相似性,用户对测量结果进行评价,根据测量结果和用户的评价动态构造一个神经网络,在神经网络反复学习过程中将图像分到不同分类中。在图像检索阶段,同样需要提取查询图像的特征。用已构造的神经网络识别查询图像所属的分类,用户用不同的标记来区分与查询图像相似的或不相似的图像,对显示结果进行评价。系统对神经网络进行动态修改,从而学习用户的反馈结果。用户的反馈过程可不断重复,直到用户对结果满意或结果不能得到进一步改善为止。由于在每次检索过程中,动态构造神经网络,检索结果被“记忆”在神经网络的神经元中,下一次检索时利用历史检索信息,可提高检索的效率。

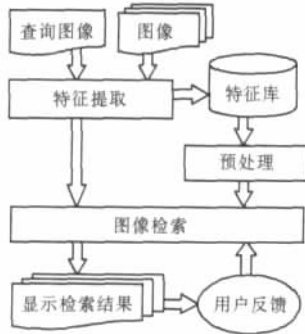


图1 图像检索系统总体结构

2.2 图像库预处理

由于图像库中图像间存在着一定的相似性,可以利用这一特点,事先将图像库中相似的图像分成一类,差异较大的图像分在不同的类中,这样在图像首次检索时可以提高检索的效率。本系统采用构造神经网络的方法对图像进行分类。神经网络学习方式分为监督学习、非监督学习和强化学习^[9]。在数据量大的图像数据库中,采用强化学习方式,即可提高分类速度,又可在一定程度上将用户的语义包含在分类中。用户对系统输出结果进行评价(奖或惩),学习系统通过强化那些受奖励的动作来改善自身性能。本系统采用此种方式。在图像预处理阶段,图像分类问题归结为图像集合的划分问题,神经网络的构造过程就是寻找集合的划分过程,将图像特征集合划分成不同的类。

在M-P神经元模型中,将神经元看成是一个具有n个输入和1个输出的处理单元,神经元的输入输出关系如下^[9]:

$$v = w^T x - \theta \quad y = \varphi(v) \tag{1}$$

其中 $x = (x_1, x_2, \dots, x_n)$ 是输入向量, $w = (w_1, w_2, \dots, w_n)$ 是向量的权值, θ 是阈值。

$$\varphi(v) = \begin{cases} 1, & v \geq 0 \\ 0, & v < 0 \end{cases} \tag{2}$$

从几何意义上讲,判决边界 $w^T x - \theta = 0$ 对应于n维空间的一个超平面,n维特征空间被分为两类,分别对应 $w^T x - \theta \geq 0$ 和 $w^T x - \theta < 0$ 。

“球面邻域”模型是对M-P神经元的另一种几何解释^[9]。当所有样本都处于同一个球面上时(若样本空间不满足这一假定,可对其做相应变换,将它映射到n+1维空间中的n维超球面上),判决边界是超平面与球面的交集,它是一个圆,该圆将整个球面分为两部分,其中 $w^T x - \theta > 0$ 空间与球面的交集称作“球面邻域”,此时对应M-P神经元输出为1,否则M-P神经元输出为0。本文将“球面邻域”的思想应用于图像检索,并在此基础上给出用于图像检索的神经网络构造方法。

设图像库为 $I = \{f_i | i = 1, 2, \dots, M\}$, 其中M为图像库中图像数

目 f_i 为图像i的特征向量,可表示为 $f_i = \langle d_{i1}, d_{i2}, \dots, d_{in} \rangle$, n是特征向量的维数。这样,一幅图像映射到n维特征空间中的一个点,图像之间的相似性度量可通过特征空间点之间的距离实现。

神经网络构造问题描述为: 设 $S = \{S^q | q = 1, 2, \dots, m\}$ 为图像集合I的一个划分,m为子集个数。求一个神经网络Net,对于所有的输入 f_i , 如果 $f_i \in S^q$, 那么目标函数为 $y^q = (y_1, y_2, \dots, y_k, \dots, y_m)$

$$y_k = \begin{cases} 1, & k=q \\ 0, & \text{其他} \end{cases} \quad q = 1, 2, \dots, m \tag{3}$$

集合划分方法的主要思路是: 随机从图像库中选取一幅图像q作为查询图像,在图像库中检索和图像q相似的图像,采用相关反馈的方法,可以由用户对检索结果进行评价。例如,用户可标记输出图像与图像q“属于同一类”或“不属于同一类”,这样在一定程度上将用户的语义包含在分类中。定义反馈函数为:

$$F(f_q, f_p) = \begin{cases} 1, & \text{如果“属于同一类”} \\ 0, & \text{如果“不属于同一类”} \end{cases} \tag{4}$$

其中 $F(f_q, f_p)$ 为图像相对于图像的标记反馈值。

假设用户标记的图像都具有一定的语义相关性,并设用户标记 $F(f_q, f_p) = 1$ 的图像p为 f_p^+ 。为了描述用户标记的图像以及与用户标记图像相关的图像在特征空间的分布情况,求一组集合 $S_1, S_2, \dots, S_k, S^q = \bigcup_{i=1}^k S_i$, 使得 f_p^+ 以及与 f_p^+ 相关的图像属于这组集合中的一个,且唯一的一个。即用一组 S_1, S_2, \dots, S_k 包含相关的图像,不相关的图像被排除在外。图2是图像特征在二维空间投影的示意图。

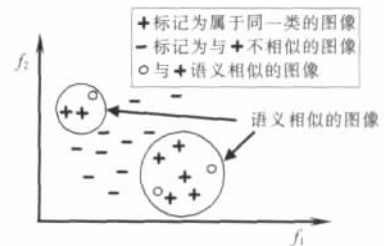


图2 神经元的构造

由于每次检索时用户标记的都是在特征空间中距离查询图像q较近的图像,虽然用户标记的图像有限,但却能够将与图像q“属于同一类”的图像和与图像q“不属于同一类”的图像用 S_1, S_2, \dots, S_k 在特征空间中分为两类。将属于 S^q 的图像删除,对图像库中剩余图像采用同样方法分类,直到图像库中图像为空为止。神经网络结构如图3所示。

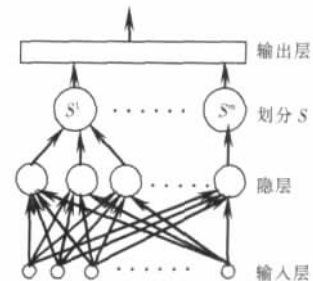


图3 神经网络结构

2.3 基于历史检索信息的交互检索

系统中的反馈历史信息包含大量的用户对数据库中图像

语义性关系评价,采用适当的形式记录并合理的利用这些信息,将明显提高系统检索的效率和效果。利用系统积累的用户反馈信息进行新一次的检索,前提条件是对用户给出的相关性评价信息进行存储和积累。本文采用动态神经网络构造方法记录每一次检索的结果,随着检索次数的增多,检索历史信息不断积累,为再一次的检索提供参照。

在预处理阶段,已经得到了图像集合 I 的一个划分 $S = \{S^1, S^2, \dots, S^m\}$, 图像被分成 m 类, 对应集合 S^1, S^2, \dots, S^m 。图像检索就是在此划分的基础上通过动态构造神经网络, 进一步寻找图像集合 I 的覆盖过程。

寻找集合覆盖的方法与图像库预处理阶段的描述基本相似, 所不同的是, 每一次检索是在图像集合划分和历史检索信息的基础上寻找查询图像 q 的一组集合, 使与 q 相关的图像属于这组集合中, 并形成图像集合新的覆盖。设图像库预处理后得到图像集合 I 的划分 $S = \{S^1, S^2, \dots, S^m\}$, 若前 n 次检索形成的图像集合 I 的覆盖为 $S \leftarrow S \{S^{n+1}, \dots, S^{m+k}, \dots, S^{m+n}\}$, 对于第 $n+1$ 次检索时, 若查询图像为 f_q , 构造的一组集合为 S_1, S_2, \dots, S_k , $S^{m+n+1} = \bigcup_{i=1}^k S_i$, 则图像集合 I 新的覆盖为 $S \leftarrow S \{S^{m+n+1}\}$, 此时 $S = \{S^1, S^2, \dots, S^m, S^{m+1}, \dots, S^{m+n}, S^{m+n+1}\}$ 。这样, 在每次检索过程中, 用户给出的图像相关性评价信息不断积累, 并存储在神经网络中, 从而提高下一次检索的效率。

神经网络的输出设计描述为: 以图像库中每一图像的特征向量 f_i 为输入, 求神经网络的输出过程。对于第 n 次检索, 若查询图像为 f_q , 构造的一组集合为 S_1, S_2, \dots, S_k , 如果 $\exists S_j$, 其中 $j = 1, 2, \dots, k$, 使输入图像 f_i 满足 $f_i \in S_j$, 即 f_i 与 f_q 属于相同的类别, 则先输出显示与 f_q 属于相同类别的所有 f_i ; 如果 $\exists S_j$, 使输入图像 f_i 满足 $f_i \in S_j$, 则计算 f_i 与各 S_j 的最小距离, 按从小到大排序, 与 f_q 不属于相同类别的图像后输出显示。

3 神经元构造算法

3.1 图像库预处理算法

设图像库为 $I = \{f_i | i = 1, 2, \dots, M\}$ 。用户只需对输出结果的前 t 幅图像进行标记, 假设用户标记“属于同一类”的图像集合为 $IN = \{f_1^+, f_2^+, \dots, f_n^+\}$, 在前 t 幅图像中, 用户标记“不属于同一类”或未标记的图像集合为 $EX = \{f_1^-, f_2^-, \dots, f_m^-\}$ 。

对于 $\forall f_i^+ \in IN$, 设置标志 tag^+ , 如果 tag^+ 为 true, 表示 f_i^+ 已被划分, 如果 tag^+ 为 false, 表示 f_i^+ 未被划分。

步骤 1 初始化, $m \leftarrow 0$, m 表示总分类数;

步骤 2 $k \leftarrow 0$, k 表示同一类中神经元数目。根据用户的标记, 输出显示的前 t 幅图像分别属于集合 IN 和 EX , 对集合 IN 中所有 f_i^+ 置 $tag^+ \leftarrow false$;

步骤 3 如果 $\exists f_i^+, f_j^+ \in IN$, $tag^+ = false$, 则计算 f_i^+ 到所有 f_j^+ 的最小距离 d_1 , 其中 $f_j^+ \in EX$, 以及 f_i^+ 到所有 f_j^+ 的最大距离 d_2 , 其中 $f_j^+ \in IN$, $tag^+ = false$, 否则, 转步骤 9;

步骤 4 如果 $d_1 < d_2$, 则 $r \leftarrow (d_1 + d_2) / 2$; 否则, $r = \beta \cdot d_1$, $\beta < 1$, $\beta \in R$;

步骤 5 在图像特征空间中, 以 f_i^+ 为中心, 对与 f_i^+ 的距离 $D(f_i^+, f_j^+) < r$ 的所有图像 f_j^+ , 增加一个集合 (C_k, R_k) , 其中 $C_k = f_i^+$, $R_k = r$, 使 $f_j^+ \in (C_k, R_k)$, 即集合 (C_k, R_k) 中所有图像 f_j^+ 满足 $D(f_i^+, f_j^+) < r$ 。以这种方式构造一个神经元;

步骤 6 对于 $\forall f_i^+$, 如果 $f_i^+ \in (C_k, R_k)$, 那么 $tag^+ \leftarrow true$;

步骤 7 对于 $\forall f_i$, 若 $f_i \in D(f_i, C_k) \cap R_k$, 则把 f_i 加入集合 (C_k, R_k) 中, 使神经元 (C_k, R_k) 包含尽量多的图像;

步骤 8 $k \leftarrow k + 1$, 转步骤 3, 继续构造神经元;

步骤 9 输出显示集合 (C_k, R_k) 中所有图像, 如果 $\exists f_i \in I$, 不属于任何 (C_k, R_k) , 则计算 f_i 与各 (C_k, R_k) 的最小距离, 按从小到大输出显示图像。

步骤 10 如果用户对结果不满意, 则标记显示的图像, 并反馈给系统, 转步骤 2;

步骤 11 $m \leftarrow m + 1$, $S^m = \bigcup_{j=0}^{k-1} (C_j, R_j)$, 集合 S^m 中图像属同一分类。从 I 中删除 S^m 中的图像;

步骤 12 如果图像库 I 为空, 转步骤 13; 否则, 转步骤 2, 即对 I 中余下的图像, 采用同样的方法继续分类。

步骤 13 $S = \{S^j | j = 1, 2, \dots, m\}$, S 为图像集合的划分, 也就是构造的神经网络。

3.2 基于历史检索信息的交互检索算法

求第 n 次检索, 查询图像为 f_q 时, 图像集合的覆盖 $S = \{S^1, S^2, \dots, S^m, S^{m+1}, \dots, S^{m+n-1}, S^{m+n}\}$ 。同样用户只需对输出结果的前 t 幅图像进行标记, 并假设用户标记与 f_q “属于同一类”的图像集合为 $IN = \{f_1^+, f_2^+, \dots, f_n^+\}$, 在前 t 幅图像中, 用户标记与 f_q “不属于同一类”或未标记的图像集合为 $EX = \{f_1^-, f_2^-, \dots, f_m^-\}$ 。

步骤 1 对查询 f_q , 用户在已经建立的覆盖 $S = \{S^1, S^2, \dots, S^m, S^{m+1}, \dots, S^{m+n-1}\}$ 中选取一个子集 S^p , 其中 $p \in \{1, 2, \dots, m, m+1, \dots, m+n-1\}$ 。如果子集 S^p 满足用户要求, 则转步骤 4, 算法结束, 否则, 继续下一步;

步骤 2 采用与图像库预处理同样的方法, 构造和查询图像 f_q 相似的图像集合 (C_k, R_k) ;

步骤 3 $S^{m+n} = \bigcup_{j=0}^{k-1} (C_j, R_j)$, 集合 S^{m+n} 中的图像在语义上与查询图像 f_q 相关。 $S \leftarrow S \{S^{m+n}\}$

步骤 4 算法结束。

4 性能评价

本系统使用的图像特征为颜色直方图和基于 DCT 变换的纹理特征。在图像颜色特征上使用 $u = R / (R + G + B)$ 和 $v = G / (R + G + B)$ 变换的颜色空间, 并在 u, v 上各取 16 个 bin, 构成 32 维颜色向量。对图像进行 3 次 DCT 变换得到 10 个子段, 取每个子段的标准差, 合成后得到一个 10 维向量, 即图像库中每一个图像用一个 42 维的向量描述。

相似性测量采用欧式距离公式, 即

$$D(f_i, f_j) = \left[\sum_{k=1}^n (d_k - d_{jk})^2 \right]^{1/2} \quad (6)$$

其中 $D(f_i, f_j)$ 表示第 i 幅图像与第 j 幅图像之间的距离。

实验系统在 Pentium 的微机采用 C++ 语言实现, 图像库中存有 2 000 幅图像, 图像内容涵盖人物、自然风景、动物、建筑、卡通画、食物等共 20 个类别, 每个类别由 100 幅语义上相近的图像组成。图像的特征值事先在离线状态下取得, 存放在 SQL Server 数据库中。

为了进行测试, 从图像库中选出 10 个类别作为测试集合。在每个测试集类别上随机选出 10 幅图像作为分类查询图像, 共选出 100 个查询图像。构造神经网络时, 参数 β 取 0.95。采用 p, r, σ 作为系统性能评价参数^[4]。

设 ID_j [1,10] 表示图像 j 所在的类别标号, $1 \leq j \leq 1000$ 。对查询图像 i , 按照 3.1 节中的算法, 在特征空间中构造一组集合 (C_k, R_k) , 使集合中的图像与图像 i 在语义上相似。对测试集中所有图像 j , 按神经网络的输出设计方法排序后输出显示, $r(i, j)$ 表示图像 j 相对于图像 i 所在的排序位置。定义图像 i 的查准率为:

$$p(i) = \frac{1}{100} \sum_{r(i, j)} 1 \quad (6)$$

$p(i)$ 表示排序后的序列中, 头 100 幅图像与图像 i 属于同一类型的百分率。

$$r(i) = \frac{1}{100} \sum_{ID_j = ID_i} r(i, j) \quad (7)$$

$$\sigma(i) = \sqrt{\frac{1}{100} \sum_{ID_j = ID_i} [r(i, j) - r(i)]^2} \quad (8)$$

$r(i)$ 表示检索结果中所有相似图像的排序中值, $\sigma(i)$ 表示标准差。

总平均查准率 p , 总平均排序中值 r , 总平均标准差 σ 定义为:

$$p = \frac{1}{100} \sum_{i=1}^{100} p(i) \quad (9)$$

$$r = \frac{1}{100} \sum_{i=1}^{100} r(i) \quad (10)$$

$$\sigma = \frac{1}{100} \sum_{i=1}^{100} \sigma(i) \quad (11)$$

对于图像分类学习, 好的性能是以图像 i 构造的一组集合 (C_k, R_k) 中所包含的图像 j 在序列中的序号 $r(i, j)$ 较低。当 p 取最大值 1, r 取最小值 50.5 时, 系统性能最好, 这时, 对于类别 k 中任何一个查询图像 i , 头 100 个检索结果图像都属于类别 k 。

图 4、图 5、图 6 显示了总的 p, r, σ 值。经过 3 次反馈后, p 从 26.1% 提高到 63.12%, r 从 474 降到 195.5, σ 从 311.3 降到 165.5。

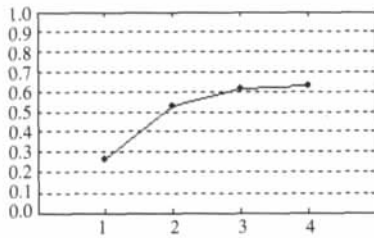


图 4 总平均查准率 p

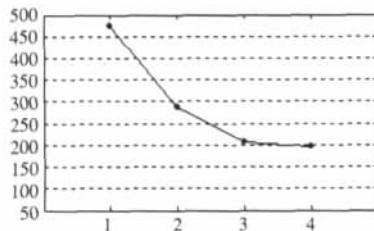


图 5 总平均排序中值 r

基于历史检索信息的图像检索方法对每类图像各进行 50 次检索, 每次检索进行 2 次反馈。10 类检索的评价结果在图 7 中给出。从图中看出, 当进行了 10 次检索后, 系统积累了 10 次历史检索信息, 系统的总 p 值得到了明显提高。

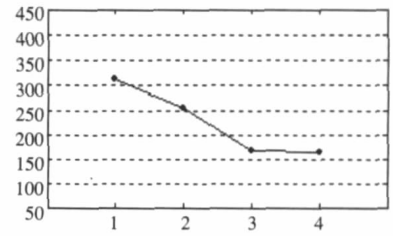


图 6 总平均标准差 σ

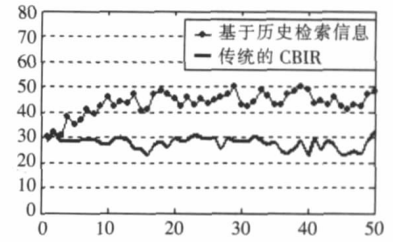


图 7 总平均精度

5 结论

本文提出了基于神经网络的交互式图像检索方法, 从实验看到, 系统根据用户的反馈信息动态调整神经网络中神经元数目, 使其逼近同类样本在特征空间的分布, 提高图像间语义的描述能力, 使分类效果得到改善。本次检索结果被“记忆”在神经网络中, 作为下次检索的历史信息, 提高了再次检索的效率, 从整体上改善了传统基于内容图像检索系统的性能。

(收稿日期 2007 年 4 月)

参考文献:

- [1] Chen Y, Wang J Z, Krowetz R. CLUE: cluster-based retrieval of images by unsupervised learning[J]. IEEE Trans Image Proc, 2005, 14(8): 1187-1201.
- [2] Chapelle O, Haffner P, Vapnik V. Support vector machines for histogram-based image classification[J]. IEEE Trans on neural networks, 1999, 10(5): 1057-1064.
- [3] Li Jia, Wang J Z, Wiederhold G. IRM: integrated region matching for image retrieval[J]. Proc ACM Multimedia, 2000: 147-156.
- [4] Wang J Z, Li Jia, Wiederhold G. SIMPLICITY: semantics-sensitive integrated matching for picture libraries[J]. IEEE Trans on Pattern Anal Mach Intell, 2001, 23(9): 947-963.
- [5] Huang Y-P, Chang T-W, Huang C-Z. A fuzzy feature clustering with relevance feedback approach to content-based image retrieval [C]// International Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2003: 57-62.
- [6] 周明全, 韦娜, 耿国华. 交互信息理论及改进的颜色量化方法在图像检索中的应用研究[J]. 小型微型计算机系统, 2006, 25(7): 181-184.
- [7] Wang Bing, Zhang Xin, Li Na. Relevance feedback technique for content-based image retrieval using neural network learning [C]// International Conference on Machine Learning and Cybernetics, 2006, 6: 3692-3696.
- [8] 张磊. 基于内容的图像检索中人机协同问题的研究[D]. 北京: 清华大学计算机科学与技术系, 2001.
- [9] 沈世铨. 神经网络系统理论及其应用[M]. 北京: 科学出版社, 2000: 11-12.
- [10] 章毓晋. 基于内容的视觉信息检索[M]. 北京: 科学出版社, 2003: 50-53.

小波域上分形噪声的一种自适应 Wiener 滤波方法

李 强, 江巨浪

LI Qiang, JIANG Ju-lang

安庆师范学院 物理系, 安徽 安庆 246011

Department of Physics, Anqing Teachers College, Anqing, Anhui 246011, China

E-mail: leelyq@126.com

LI Qiang, JIANG Ju-lang. 1/f fractal noise attenuation using adaptive Wiener filter. Computer Engineering and Applications, 2007, 43 (22): 63-66.

Abstract: A multiscale adaptive Wiener filter based on wavelet is proposed for signal restoration embedded in 1/f fractal noise. First the signal with 1/f fractal noise is transformed into multiscale subsystems in time-scale domain. Some nonstationary properties of fractal signals are attenuated in each subband by wavelet multiresolution decomposition so that the Wiener filter bank can be applied to estimate the multiscale input signals. Then the estimated multiscale input signals are synthesized to obtain the estimated signal. Some simulation examples are given for testing the performance of the proposed algorithm and output signal to noise ratio is improved evidently.

Key words: Wiener filter, wavelet transform, 1/f fractal noise, self-similar

摘 要: 针对淹没在 1/f 分形噪声中的有用信号恢复问题, 提出了一种基于小波变换与 Wiener 滤波的多尺度自适应滤波算法。首先将带有 1/f 分形噪声的信号分解成多尺度的子带信号, 通过小波变换对 1/f 分形噪声的白化作用, 消除了 1/f 分形噪声的自相似性和长程相关性。然后在小波域内, 利用自适应 Wiener 滤波实现了噪声和有用信号的分离, 估计出了各子带中的有用信号。最后进行小波重构, 较好地恢复出淹没在 1/f 分形噪声中的有用信号。仿真实验表明, 使用多尺度自适应 Wiener 滤波器能有效地抑制分形噪声, 显著地提高信噪比。

关键词: Wiener 滤波; 小波变换; 1/f 分形噪声; 自相似性

文章编号: 1002-8331(2007)22-0063-04 文献标识码: A 中图分类号: TP391

1 引言

1/f 分形信号是分形信号中重要的一类, 广泛地存在于各种自然现象和社会现象中, 如心率及脑波的波动、电子元器件中的噪声、通信信道上的突发误差等。它具有不平稳、自相似、长程相关性、1/f 类型的谱密度等特性^[1,2]。近年来 1/f 分形信号的处理成为信号检测与估计、信号处理及图像处理的研究热点, 其在生物医电、语音、图像、雷达信号、网络通信等众多领域有着广泛的应用^[3,4]。从 1/f 分形噪声中提取、估计、检测有用信号是信号处理领域经常遇到的一类问题。由于 1/f 分形信号的特性, 传统的信号检测方法不能有效地去除其自相似性及长程相关性, 而小波变换具有多尺度分析能力, 成为分析此类信号的有力工具^[5,6]。本文针对以 1/f 类分形噪声为背景噪声的有用信号估计问题, 提出一种基于小波变换域上的自适应 Wiener 滤波算法。仿真实验表明, 该算法能有效地抑制 1/f 分形噪声, 显著地提高了信噪比。

2 1/f 分形信号的小波变换分析

2.1 1/f 分形信号的统计特性

一般地说, 分形信号 $x(t)$ 是这样一随机过程, 它的重要特

征是在时间尺度上表现出局部自相似性^[2]:

$$x(t) \stackrel{P}{=} a^H x(at) \quad (1)$$

其中 a 为一常数, H 为相似参数, $\stackrel{P}{=}$ 表示统计意义上的相等。对于满足上式的分形信号, 其均值、自相关函数统计特征均满足自相似性:

$$M_x(t) = E(x(t)) = a^H M_x(at) \quad (2)$$

$$R_x(t, s) = E(x(t)x(s)) = a^{2H} R_x(at, as) \quad (3)$$

如果 $x(t)$ 是广义平稳的, 则有:

$$R_x(\tau) = a^{2H} R_x(a\tau) \quad (4)$$

其中 $\tau = s - t$ 。于是可得功率谱的自相似性:

$$S_x(\omega) = a^{2H+1} S_x(a\omega) \quad (5)$$

满足上面功率谱自相似关系的功率谱具有下面的形式:

$$S_x(\omega) = \frac{\sigma_x^2}{|\omega|^{2H+1}} \quad (6)$$

其中 $\sigma_x > 0$, $\gamma = 2H+1$ 称为分形指数。在一定频率范围内, 如果具有上述 $\frac{1}{|\omega|^\gamma}$ 形式的功率谱密度的随机过程称为 1/f 类分形信号。

对 $S_x(\omega)$ 作傅里叶变换反变换可得 $R_x(\tau) \propto |\tau|^{-2H}$, 也即 $R_x(\tau) \propto |\tau|^{-2H}$ 。

作者简介: 李强 (1971-) 男, 讲师, 主要研究内容为信号检测与处理技术; 江巨浪 (1967-) 男, 副教授, 博士, 主要研究内容为计算机图像处理。

随 按乘幂律衰减,其衰减速度比指数律衰减慢得多,这说明分形信号在时间上具有长程相关性。

为了研究 1/f 分形信号, Mandelbrot 等人对传统的布朗运动进行推广,提出了分数布朗运动 $B_H(t)$ (记作 $fBm, 0 < H < 1$), 用来描述具有 1/f 类型谱密度的长程相关信号。因其在数学处理上的简便性而成为 1/f 分形信号的一个经典模型^[3,4]:

$$B_H(t) = \frac{1}{\Gamma(H+0.5)} \left[\int_0^t (t-s)^{H-0.5} |s|^{H-0.5} w(s) ds + \int_0^1 |t-s|^{H-0.5} w(s) ds \right] \quad (7)$$

其中 $w(s)$ 为一平稳高斯白噪声。 $B_H(t)$ 为一零均值的高斯非平稳过程,其自相关函数为:

$$R_x(t, s) = \frac{\sigma^2}{2} [|t|^{2H} + |s|^{2H} - |t-s|^{2H}] \quad (8)$$

其中 $\rho^2 = \Gamma(1-2H) \frac{\cos H}{H}$ 。该式表明 fBm 是非平稳的随机过程,而且具有长程相关性。对于这一非平稳的随机过程,也可得其经验功率谱为^[4,5]:

$$S_{B_H}(w) = \frac{\sigma^2}{|w|^{2H+1}} \quad (9)$$

可见 fBm 也可看作 1/f 分形信号。

2.2 1/f 分形信号的小波分析

传统的信号检测理论如相关检测、匹配滤波等大都假定背景噪声是平稳的白噪声。于是要从具有局部自相似及长程相关性的分形噪声中检测出有用信号,由于传统的一些分析方法不能有效地去除它的自相似及长程相关性,在处理上遇到了困难。正是由于数学上一直缺乏有力的分析工具,使得该类信号的分析和应用受到很大的限制。近年来成熟起来的小波变换理论,同时具有时变分析和尺度分析的功能,这为人们研究具有长程相关性的 1/f 分形信号提供了一个有力的工具。

小波变换是把待分析的信号用基本小波 $\psi(t)$ 的尺度伸缩和平移作处理,当尺度按二进制变化时,基本小波 $\psi(t)$ 的尺度伸缩和平移定义为^[1]:

$$\psi_j^k(t) = 2^{-\frac{j}{2}} \psi(2^{-j}t - k) \quad (10)$$

式中 2^j 是尺度因子, k 是平移。此时对于分形信号 $x(t)$ 的离散小波变换(DWT)系数可以表示为^[1]:

$$d_j^k = x(t) \psi_j^k(t) = 2^{-\frac{j}{2}} \int_{-\infty}^{\infty} x(t) \psi(2^{-j}t - k) dt \quad (11)$$

式中, \int 符号表示作内积运算, j 是尺度序号, k 是时间序号。

Wornell^[4,5]等详细地研究了 1/f 分形信号小波变换系数 d_j^k 的特性。并证明如果 $0 < \gamma < 2R$, 对任意两个小波系数 $d_{j_1}^{k_1}, d_{j_2}^{k_2}$ 间具有如下相关特性:

$$E(d_{j_1}^{k_1} d_{j_2}^{k_2}) = \frac{E(d_{j_1}^{k_1} d_{j_2}^{k_2})}{\sqrt{\text{Var}(d_{j_1}^{k_1}) \text{Var}(d_{j_2}^{k_2})}} \sim O(|2^{j_1} k_1 - 2^{j_2} k_2|^{-2R-\gamma}) \quad (12)$$

其中 R 为所使用的小波函数的正规性,并且任意两尺度上系数的相关性总量相同,即当尺度放大或缩小时,小波系数的二阶统计特性不变^[2]。由此可见 1/f 分形信号小波变换系数有下列特性:

(1) 在同一尺度下 d_j^k 是近似的平稳的随机序列,且其相关系数的衰减速度远快于分形信号本身的自相关函数的衰减速度,

因此可以近似认为是白噪声。

(2) 在不同尺度间小波变换系数的方差沿尺度指数变化,并且有下面的关系式:

$$\text{Var}(d_j^k) = \sigma^2 2^j \quad (13)$$

(3) 对小波函数 $\psi(t)$,其正规性愈大,则同一尺度下及不同尺度间的小波变换系数相关性衰减愈快。

这些特性给 1/f 分形信号的处理带来了很大方便,一方面它提供了一种通过小波变换生成 1/f 信号的方法;另一方面它说明小波变换对自相似的 1/f 信号有很强的去相关作用,小波变换将 1/f 噪声时间域的自相似性变换到小波域的尺度上来,在同一尺度下其系数基本上不相关。因此小波变换可以用作 1/f 噪声的白化滤波器,去除其时间上的自相似性及长程相关性,使得 1/f 噪声中信号估计易于在小波域上实现。

3 1/f 分形噪声的自适应 Wiener 滤波

在分形信号处理中,以前的工作较多集中在分形噪声的参数估计上,但估计分形噪声中信号的工作并不多^[3,5]。将以估计淹没在分形噪声中的信号为研究对象,假设得到的观测数据为:

$$y(n) = s(n) + w(n) \quad (14)$$

其中 $w(n)$ 是 1/f 分形噪声, $s(n)$ 为有用信号,与 $w(n)$ 互相独立。

由于 1/f 分形噪声的非平稳性,使得直接实施平稳信号的分析方法不能取得好的效果。由上章分析,小波变换是分析 1/f 过程的良好工具。1/f 过程的小波系数是平稳的,近似不相关的,可看作近似高斯白噪声。所以将 1/f 分形噪声经小波变换,以上问题就可简化为处理平稳白噪声的问题。

首先将 $y(n)$ 做离散小波变换,得:

$$y_j(n) = s_j(n) + w_j(n), j=1, \dots, J \quad (15)$$

$y_j(n)$ 是第 j 层分解产生的小波系数。其中 1/f 的小波变换系数 $w_j(n)$ 已近似成为白噪声序列。它的方差满足式(13)。

现在主要问题是在各个分解层从白噪声中估计出有用信号。本文采用一种自适应 Wiener 滤波器来进行最佳估计。

小波变换对信号具有近似解相关的作用,并将信号能量集中在一个相对较小的子空间上。对很多信号,小波变换可以作为一个近似的 KL 基,对于得到的小波系数,在一定意义上可以认为是相互独立的,所以在小波域上同样可以采用对角形式的 Wiener 滤波器,以获得最佳结果。对角化的 Wiener 滤波器为:

$$\hat{s}_j(n) = \frac{\hat{s}_j^2(n)}{\hat{s}_j^2(n) + w_j^2} y_j(n) \quad (16)$$

式中 w_j^2 为 $w_j(n)$ 的方差。因为 $s_j(n)$ 是未知的,所以只能使用其估计值 $\hat{s}_j(n)$,这样就得到经验 Wiener 滤波器:

$$\hat{s}_j(n) = \frac{\hat{s}_j^2(n)}{\hat{s}_j^2(n) + w_j^2} y_j(n) \quad (17)$$

由公式(17)可知,为设计小波域上的 Wiener 滤波器,需从输入数据中估计 $\hat{s}_j(n)$ 和 w_j^2 。准确估计 $\hat{s}_j(n)$ 比较困难,因为通过小波系数仅可估计出噪声方差 w_j^2 。对正交小波分解来说,在不同的尺度,其系数可以看作是独立,并且信号的支撑区间随着尺度的变化而变化。于是利用各尺度系数的去相关以及局部统计特性可以设计出小波域局部自适应 Wiener 滤波器。

对于尺度系数可以采用下面的滤波器：

$$\hat{s}_j(n) = m_{y_j}(n) + \frac{\hat{s}_j(n)}{s_j(n) + w_j} (y_j(n) - m_{y_j}(n)) \quad (18)$$

其中：

$$m_{y_j}(n) = \frac{1}{2p+1} \sum_{i=n-p}^{n+p} y_j(i) \quad s_j(n) = \max(y_j^2(n) - w_j^2, 0)$$

其中 s_j^2 小波系数的局部方差 $2p+1$ 为滤波器窗口的宽度 m_{y_j} 为小波系数的局部均值。

由于小波系数的局部均值接近于零，所以小波系数可使用下面的滤波器：

$$\hat{s}_j(n) = \frac{\tilde{s}_j^2(n)}{\tilde{s}_j^2(n) + w_j} y_j(n) \quad (19)$$

$$\text{其中 } \tilde{s}_j(n) = \max(\hat{s}_j^2(n) - w_j^2, 0) \quad \tilde{s}_w^2(n) = \frac{1}{2p+1} \sum_{i=n-p}^{n+p} y_j^2(i)$$

将 Wiener 滤波的结果 $\hat{s}_j(n)$ 进行小波逆变换，就得到了有用信号的估计 $\hat{s}(n)$ 。至此，设计出了分形噪声中信号恢复的多尺度 Wiener 滤波器。

基于离散小波变换的 $1/f$ 噪声多尺度 Wiener 滤波方法可归纳为：

- ① 根据实际情况，应用离散小波变换把数据分解在若干个尺度上。
- ② 在各个尺度上分别进行 Wiener 滤波，得到各个尺度上的滤波结果。
- ③ 将各个尺度上的滤波结果，应用离散小波变换逆变换，得到原信号滤波后的结果。

4 仿真计算

为了验证本文算法的性能，进行了仿真计算。使用不同 H 指数的分数布朗运动作为 $1/f$ 分形噪声，并用一个 AR(3) 序列作为原始信号：

$$s(n) = 1.5s(n-1) - 0.75s(n-2) - 0.125s(n-3) + u(n) \quad (20)$$

$u(n)$ 为一白噪声。

图 1 是 AR(3) 序列实现，图 2 为 $H=0.4$ 时 $1/f$ 分形噪声的一个实现，数据长度为 1024 个采样点。图 3 与图 4 中从上至下依次为白噪声、图 2 中的 $1/f$ 分形信号及其小波分解系数序列的自相关函数 (ACF) 估计及功率谱 (PSD) 估计。从这两图中可看出 $1/f$ 分形信号的长程相关性及其的 $1/f$ 谱性质，通过对它的小波分解系数的序列相关函数估计及功率谱估计与白噪声的相应估计比较，可以看出 $1/f$ 分形信号的小波分解系数序列近似为白噪声。图 5 与图 6 分别是观测数据及采用本文算法滤波后的数据波形和估计误差，仿真实验时采用 Daubechies4 小波，分解尺度为 $J=4$ 。从结果上可看出 Wiener 滤波方法获得了较好的效果，滤波后信号的信噪比提高了 20 dB 左右。

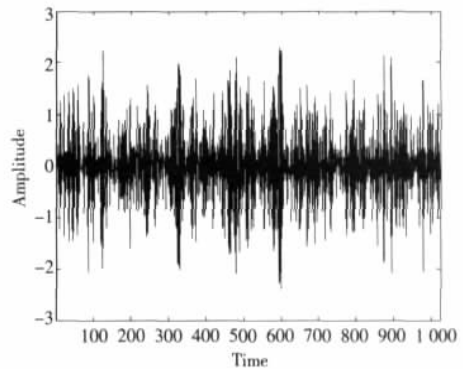


图 1 原始数据

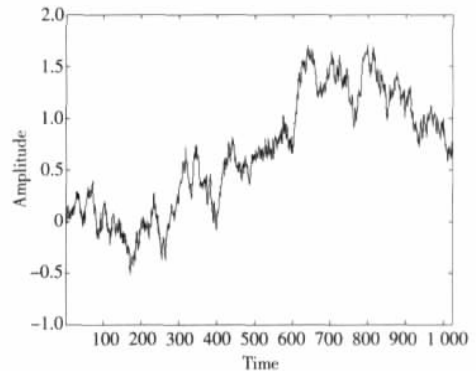


图 2 分形噪声 (H=0.4)

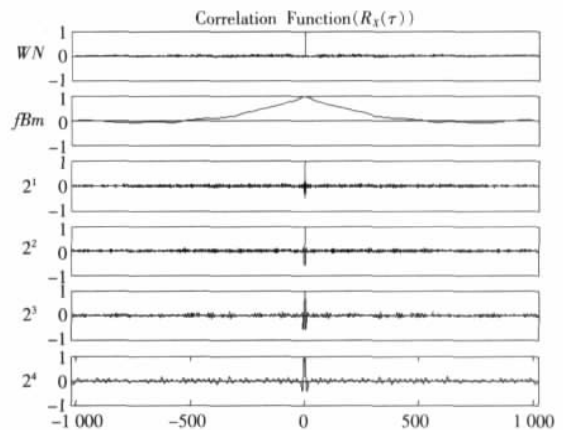


图 3 相关函数的比较

为了验证本文滤波器的性能，使用信噪比 (SNR) 及均方误差 (MSE) 两个量来评估。表 1 给出了在相同 $s(n)$ ，而加入不同 H、信噪比为 0 dB 的 $1/f$ 分形噪声时，在不同窗口阶次 p 及分解尺度 J 情况下多尺度 Wiener 滤波器的性能统计，仿真时使用的是 Haar 小波。由表 1 可以看出，对不同 H 的 $1/f$ 噪声，本文的算法都具有较强的噪声抑制能力，滤波后的信噪比得到了较大的提高。并且随着滤波器的窗口阶次 p 的增加，估计结果也就越精确。对不同的分解尺度 J ，估计的结果也不尽相同。一般来说，取较大分解的尺度 J ，估计的结果也会好一点。

表 1 不同 J 及 p 情况下，Wiener 滤波器的性能

H	p=5, J=4		p=7, J=4		p=9, J=4		p=9, J=5		p=9, J=6	
	SNR	MSE	SNR	MSE	SNR	MSE	SNR	MSE	SNR	MSE
0.3	17.768 6	0.011 3	18.441 1	0.011 1	18.746 0	0.010 3	18.749 0	0.010 4	18.751 0	0.010 2
0.4	20.286 3	0.007 2	20.457 0	0.006 9	20.488 4	0.006 9	20.490 5	0.006 9	20.499 0	0.006 9
0.6	22.664 3	0.003 7	22.961 7	0.003 4	23.190 0	0.003 2	23.198 0	0.003 2	23.210 5	0.003 0
0.7	24.256 3	0.002 5	24.492 9	0.002 4	24.662 1	0.002 3	24.670 0	0.002 3	24.674 5	0.002 1

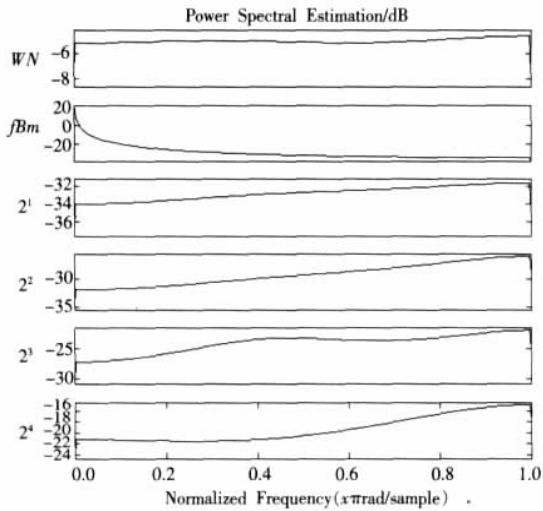


图4 功率谱的比较

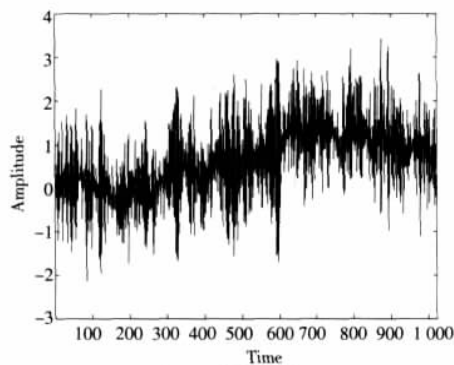


图5 观测信号

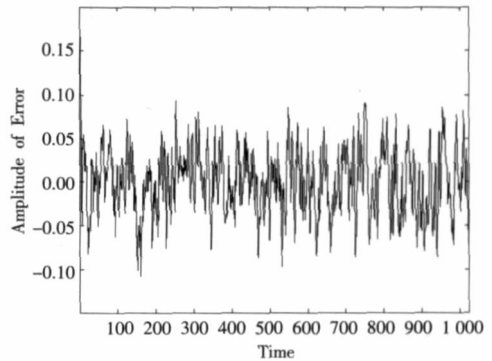


图6 滤波估计误差

性,得到信号的较好估计。自然界中有许多非平稳信号在一定尺度范围内具有分形特性,因此,本文可广泛用于分形噪声背景下的信号估计,将有助于分形噪声中有用信号的提取、估计及检测。(收稿日期 2006 年 11 月)

参考文献:

- [1] 杨福生.小波变换的工程分析与应用[M].北京:科学出版社,2000.
- [2] 阎晓红,刘贵忠.分形信号的多小波基表示[J].自然科学进展,2004,14(3):354-358.
- [3] 张淑宁,熊刚,赵惠昌,等.分形随机噪声信号处理的小波谱相关方法[J].电子学报,2005,33(7):1213-1217.
- [4] Wornell G W. Wavelet-based representations for the $1/f$ family of fractal processes[J]. Proceedings of The IEEE, 1993, 81(10):1428-1450.
- [5] Wornell G W, Oppenheim A V. Estimation of fractal signals from noisy measurements using wavelets [J]. IEEE Transactions on Signal Processing, 1992, 40(3):611-623.
- [6] 王尔丹,李晓华,沈兰荪.基于多尺度分析和分形的人群密度估计方法[J].计算机工程与应用,2005,41(29):35-38.
- [7] 宿丁,张启衡,谢盛华.一种强海杂波多目标分形分割算法[J].计算机工程与应用,2006,42(16):12-14.
- [8] 李士心,刘鲁源.小波变换域的局部自适应 Wiener 滤波器设计方法研究[J].信号处理,2003,19(2):185-187.
- [9] 郭炳庆,卢强,黎雄.一种基于信号局部统计量的自适应 L 滤波器[J].电子学报,2001,29(7):909-911.

5 结束语

根据分形噪声的小波变换系数特点,本文提出了一种多尺度 Wiener 滤波方法来估计淹没在分形噪声中的有用信号。针对每一尺度,给出了相应的系统状态方程和输出方程,从原理上阐明了其可行性,并通过仿真验证了它的有效性。计算机仿真结果表明,通过本方法,能够有效消除分形噪声的长程相关

(上接 49 页)

质来实现更快的加密。(收稿日期 2007 年 3 月)

参考文献:

- [1] Pichler F, Scharinger J. Ciphering by bernoulli shifts in finite abelian groups[C]//Proc Conference on Contributions to General Algebra Linz-Conference, 1994:465-476.
- [2] Fridrich J. Symmetric ciphers based on two-dimensional chaotic maps[J]. Int J Bifur Chaos, 1998, 8(6):1259-1264.
- [3] Chen G, Mao Y, Chui C. Symmetric image encryption scheme based on 3D chaotic cat maps[J]. Chaos, Solitons & Fractals, 2004, 21(3):749-761.
- [4] Lian S, Shun J, Wang Z. A block cipher based on a suitable use of the chaotic standard map[J]. Chaos Soliton & Fractals, 2005, 26(1):117-129.
- [5] Guan Z, Huang F, Guan W. Chaos-based image encryption algorithm[J]. Physics Letters A, 2005, 346(1/3):153-157.
- [6] Zhang L, Liao X, Wang X. An image encryption approach based on chaotic maps[J]. Chaos, Solitons & Fractals, 2005, 24(3):759-765.
- [7] Gao H, Zhang Y, Liang S, et al. A new chaotic algorithm for image encryption[J]. Chaos, Solitons & Fractals, 2006, 29(2):393-399.
- [8] Furht B, Kirovski D. Multimedia security handbook[M]. Boca Raton, Florida: CRC Press, 2005.
- [9] Maniccam S, Bourbakis N. Image and video encryption using SCAN patterns[J]. Pattern Recognition, 2004, 37(4):725-737.
- [10] Pommer A, Uhl A. Selective encryption of wavelet-packet encoded image data: efficiency and security[J]. Multimedia Systems, 2003, 9(3):279-287.
- [11] Wu C, Kuo C. Design of integrated multimedia compression and encryption systems[J]. IEEE Transaction on Multimedia, 2005, 7(5):828-839.
- [12] Cheng H, Li X. Partial encryption of compressed images and videos[J]. IEEE Transactions on Signal Processing, 2000, 48(8):2439-2451.
- [13] Nechvatal J. Report on the development of the advanced encryption standard[R]. National Institute of Standards and Technology, U.S. Dept Commerce, Oct 2000.
- [14] NIST. Federal information processing standards publication 197 [EB/OL]. <http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf>.
- [15] Scharinger J. Application of signed kolmogorov hashes to provide integrity and authenticity in Web-based software distribution[C]// Formal Methods and Tools for Computer Science, 2001:85-88.

机器人路径规划中的改进型遗传算法

唐国新 陈 雄 袁 杨

TANG Guo-xin ,CHEN Xiong ,YUAN Yang

复旦大学 信息学院 ,上海 200433

School of Information Science and Technology ,Fudan University ,Shanghai 200433 ,China

E-mail :042021044@fudan.edu.cn

TANG Guo-xin ,CHEN Xiong ,YUAN Yang.Improved genetic algorithm for robotic path planning.Computer Engineering and Applications 2007 43 (22) :67- 70.

Abstract : An improved genetic algorithm designed for robotic path planning is proposed.Crossover operator and mutation operator are specifically optimized for the application of path planning ,and two customized genetic operators are introduced as well.The evolutionary efficiency is noticeably enhanced by incorporating map information in genetic operators and better adaptive ability towards complex maps is realized by adopting the customized genetic operators.Demonstrated by simulation results ,the proposed algorithm achieves considerable improvements ,with respect to the basic genetic algorithm ,in optimal solution output rate.

Key words : genetic algorithm ,robotic path planning ,crossover operator ,mutation operator ,customized genetic operator

摘 要 提出了一种应用于机器人路径规划的改进型遗传算法。针对机器人路径规划的实际应用 ,优化设计了交叉算子和变异算子 ,引入了自定义的插入和删除两种遗传操作。通过把地图特征信息作为参与决策的已知条件来约束遗传算子的操作过程 ,提高了算法的进化效率。自定义遗传算子的使用 ,使得算法对复杂地图也表现出良好的适应能力。计算机仿真实验证明该算法在最优解输出概率方面相对于基本遗传算法有了显著提高。

关键词 遗传算法 ;机器人路径规划 ;交叉算子 ;变异算子 ;自定义遗传算子

文章编号 :1002- 8331 (2007)22- 0067- 04 文献标识码 :A 中图分类号 :TP18

1 引言

在机器人学领域 ,路径规划是最基本同时也是最重要的研究课题之一。作为机器智能的一部分 ,它的任务就是使机器人在其工作环境中能够自主的从起始点运动到目标点同时满足一定的约束条件。这些约束条件包括但不限于 : (1) 不与障碍物碰撞 ; (2) 运动路径最短 ; (3) 尽量远离障碍物 ; (4) 机器人耗能能量最小 ; (5) 运动轨迹包含某些设定的必经路点。

基于不同的基本思想 ,有多种算法出现。包括有人工势场法 (APF) 、臭虫法 (BUG) 、随机位图法 (PRM) 、快速随机树 (RRT) 、神经网络法 (NN) 等等。这些算法都在不同的方面有着各自的优势 ,然而在总体上 ,这些路径规划算法都还存在各自的不足。如计算复杂度、局部最优解、地图适应性、动态环境下有效性、全局及多目标优化能力。

随着遗传算法 (GA) 的提出 ,机器人路径规划算法也得到相应的发展。通过 GA 在路径规划中的应用^[1,2,7] ,使得机器人更加的智能 ,其运行路径也更加逼近理想的优化要求。尤其是其优秀的全局优化能力和较高的搜索效率使得 GA 成为处理复杂环境下全局优化要求的最佳算法之一。

早期的路径规划应用中 ,GA 算法通常未对路径规划问题特别设计遗传操作 ,而使用的标准遗传操作又无法反映所求问

题的特定知识 ,所以算法的进化过程很大程度上属于盲搜索状态。较低的效率使得一条最终路径的产生需要耗费相当长的时间^[2]。为了利用路径规划问题中已有的知识来加快求解速度 ,常用的方法是先构造 MAKLINK 图来进行地图建模以找出机器人可以运动的自由空间 ,然后再使用 Dijkstra 算法得出 MAKLINK 图中的最短路径 ,最后 ,再应用 GA 对所得路径进行优化 ,以获得全局最短路径。这样的方法虽然可以有效的简化 GA 需要完成的优化工作 ,但它需要完成前期非常耗时的建模过程作为运行 GA 算法的前提。所以 ,其总体计算复杂性并没有得到有效地降低。后续过程对地图建模的依赖性使得算法面对不同的地图环境时缺乏灵活性。

针对上述不足之处 ,本文采用了直接基于路点坐标值的可变长度的染色体编码方式 ,配合包含障碍物排斥子函数项的代价函数 ,有效避免了地图建模过程。同时针对路径规划问题的具体应用改进了交叉和变异两种标准遗传操作 ,并且还引入了插入和删除两种自定义遗传操作 ,大大提高了算法的搜索效率 ,实现了较高的收敛速度。计算机仿真结果证明了本方法较强的地图适应性和相对于标准遗传算法的性能提高。

本文内容安排如下 ,第 2 章介绍所提的改进型遗传算法 ,第 3 章给出算法在多种地图环境下的仿真结果 ,以及和普通遗传算法的性能对比。第 4 章作出总结。

作者简介 :唐国新 (1981-) ,男 ,在读硕士研究生 ,主要研究方向为机器人路径规划 ;陈雄 (1964-) ,男 ,博士 ,副教授 ,主要研究方向为机器学习、多机器人协作 ;袁杨 (1982-) ,女 ,在读硕士研究生 ,主要研究方向为机器人路径规划。

2 改进型遗传算法

在本文的机器人路径规划中,目标是在一幅障碍物分布已知的二维地图上寻找一条最优路径使到达目标点距离最短,同时尽可能地最大化与障碍物的距离。为了简化讨论,机器人被作为一个质点来考虑,而障碍物的边界向外扩张半个机器人最大直径。

2.1 染色体编码

机器人的运动轨迹由若干直线段构成,每段直线段是机器人运动的基本单位,于是机器人到达目标点的整个路径可表示成

$$T=l_1+l_2+\dots+l_{n-1}$$

其中 l_i 为第 i 段直线段的矢量表示,它的两个端点分别为 p_i 和 p_{i+1} ,符号“+”表示把各矢量首位相连的连接运算。以 O 表示原点,于是有

$$l_i=OP_{i+1}-OP_i$$

进而,整个机器人运动轨迹可以表述为路点矢量的顺序集合

$$T=\{OP_1,OP_2,\dots,OP_n\}$$

设 p_i 点坐标为 (x_i, y_i) ,于是算法实现时,路径 T 存储为

$$T=\{x_{i+1}+iy_1, x_{i+1}+iy_2, \dots, x_{i+1}+iy_n\}$$

其中 i 为虚数单位。于是完成了直接基于路点坐标的矢量染色体编码。在后续算法过程中,固定 (x_1, y_1) 为机器人起始点坐标, (x_n, y_n) 为目标点坐标,所以路径 T 是一个连接起点和目标点的可能解。其中 n 为变数并满足 $2 \leq n \leq N_{max}$, N_{max} 为最大路点数。

2.2 适应度函数

每条路径的优劣评价通过适应度函数来给出。本文中,以路径长度和障碍物相交程度作为评价指标,并使所求解向这些指标渐小的方向进化,所以这里的适应度函数又被称为代价函数。使各项指标在代价函数中表现出来,于是该函数构造如下:

$$F(T_k)=a_1 \sum_{i=1}^{N-1} l_i + a_2 \sum_{i=1}^{N-1} \alpha_i$$

其中 a_1, a_2 为权重系数,分别强调了不同优化指标的重要性。式中第 1 项为表示路径 T_k 的总长度, l_i 表示路径 T_k 中第 i 段直线段的长度。适应度函数中的第 2 项被称为障碍物排斥函数,表达为:

$$\alpha_i = \sum_{j=0}^M \beta_{ij}$$

其中 M 为障碍物的个数,而 β_{ij} 则是第 i 段直线与和第 j 个障碍物的排斥度。并定义为

$$\beta_{ij} = \begin{cases} \gamma + d_s & \text{, 直线段与障碍物相交时} \\ \frac{1}{d_s} \cdot d_0^2 + (-2) \cdot d_0 + d_s & \text{, 直线段距离障碍物 } d_0 < d_s \\ 0 & \text{, 直线段远离障碍物 } d_0 > d_s \end{cases}$$

其中 γ 为使直线段不与障碍物相交所要移动的最短距离, d_0 为直线段到障碍物的距离,称 d_s 为安全距离,当 $d_0 < d_s$ 后,算法将不再试图使路径进一步远离障碍物,称该线段和障碍物无排斥。

在后面的运行过程中,算法试图使代价函数最小化并认为使得该函数取得较小值的解为较优解。

2.3 遗传操作

遗传操作是决定 GA 性能的关键因素之一。本文针对路径规划问题的特点改进了标准交叉算子和变异算子,把地图信息作为参与决策的已知条件引入遗传操作以改善算法在解空间上的搜索效率。此外,除了改进标准遗传操作,本文还引入了自

定义的插入和删除操作,进一步提高算法的搜索能力。

2.3.1 交叉算子

交叉操作作用于两个操作对象,分别对它们进行随机分割,然后再重新组合从而获得两个新个体。根据分割点的数量,可分为单点交叉或是多点交叉。其原理都可用多点交叉过程加以说明,对每条路径随机选择 m 个无重复的交叉点,在交叉点之间的变量间断地相互交换,产生两条新的路径,完成交叉操作。如图 1。

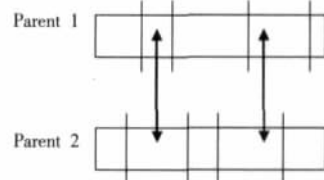


图 1 多点交叉操作示意

图 1 中,两个父个体均被 4 个分割点分为 5 部分,其中标有箭头的部分进行相互交换。

显然,由于分割点位置的不同,当交叉操作完成后,新产生的两个子代个体的染色体长度很有可能是不相等的并且相对于父代也发生变化。同时,其对应的代价函数值也将改变。所以,本文对交叉算子进行改进,在分割点个数确定后,寻找合适的分割点位置,使得新产生的子代个体之一获得尽可能低的代价函数值。

为此,将路径 T 写为

$$T=l_1+l_2+\dots+l_{n-1}$$

$$l_i=OP_{i+1}-OP_i$$

记路径上每条线段的相交函数为:

$$f(l_i) = \begin{cases} 0 & \text{, 第 } i \text{ 段直线段与所有障碍物不相交} \\ 1 & \text{, 第 } i \text{ 段直线段与障碍物相交} \end{cases}$$

从而

$$g=f(l_{i+1}) \cdot f(l_i)$$

指出出路径上各线段与障碍物相交状态的变化。当 g 为 1 时,说明 $i+1$ 点的前一线段未与障碍物相交而其后一线段与障碍物发生相交。反之,当 g 为 -1 时,说明 $i+1$ 点的前一线段与障碍物相交,而其后一线段未与障碍物相交。当 g 等于 0 时,说明在 $i+1$ 点前后的两段线段与障碍物的相交状态相同。

选择 g 为 1 时对应的变化点作为 1 号父个体的第一分割点,选择紧随该点之后使得 g 为 -1 的点作为第二分割点。对于 2 号父个体,选择过程恰好相反,选择 g 为 -1 时对应的变化点作为 2 号父个体的第一分割点,选择紧随该点之后使得 g 为 1 的变化点作为第二分割点。更多的分割点同理可得。

至此,对分割点位置的优化得以完成。但是,对于分割点数量的选择依然需要考虑。由于机器人路径规划问题自身的特点,以及本文采用的可变长度染色体编码方案,染色体长度在算法初始阶段会较长,所以采用多点交叉以改善算法搜索性能。然而,如前所述,非定点的交叉操作使得算法的进化后期染色体长度可能会变得较短。于是,在进行多点交叉操作时,同时考虑根据不同的进化阶段,不同的染色体长度设定不同的交叉点数。在算法设计中,本文采用如下分段设定:

$$crossnm = \begin{cases} 1 & 2 < clen < 5 \\ 2 & 5 < clen < 20 \\ 4 & 20 < clen < 35 \\ 6 & 35 < clen < N_{max} \end{cases}$$

其中, $crossnm$ 表示交叉点数, $clen$ 表示染色体长度, N_{max} 则为本

算法中限定最大染色体长度。

2.3.2 变异算子

变异过程中, 个体中的分量以很小的概率或步长产生转移。对于给定路径, 该操作对路径上的各路点 p_i 以一定的概率 P_{mu} 改变其坐标。

在标准变异过程中并没有对已知的地图信息加以利用。使得变异操作成为了随机搜索过程, 于是变异常常导致路径劣化。为此, 改进型变异算子优先选取和障碍物相交的线段的端点进行变异, 同时限制变异所得的路点坐标在障碍物之外, 并且使变异所得的路点新坐标满足

$$l_i^{new} = OP_{i+1} - OP_i^{new}$$

$$l_{i-1}^{new} = OP_i^{new} - OP_{i-1}$$

$$f(l_{i-1}^{new}) + f(l_i^{new}) = f(l_{i-1}) + f(l_i)$$

通过这样的约束条件保证了每次变异对路径优化的非负面效果。

2.3.3 插入算子

该算子在其所作用路径上增加路点。考虑路径上某一直线段 $l_i = OP_{i+1} - OP_i$ 与障碍物相交, 并且有端点坐标 p_i 处于障碍物外部空间。于是, 通过在 p_i 和 p_{i+1} 之间插入合适的端点 p_{i+1}^{new} , 一定可以得到 $l_i^{new} = OP_{i+1}^{new} - OP_i$ 不与障碍物相交。同理, 对于 p_{i+1} 处于障碍物外部空间时, 一定可以有 $l_{i+1}^{new} = OP_{i+1}^{new} - OP_{i+1}$ 不与障碍物相交。对于 p_i 和 p_{i+1} 均位于障碍物内部的情况, 该算子将随机生成坐标值, 满足 p_{i+1}^{new} 位于所有障碍物的外部空间。

2.3.4 删除算子

该算子在其所操作路径上记录所有位于障碍物内部空间的路点, 随机选择其中之一并予以删除。对于不和障碍物相交的路径, 该算子则在其全体路点中随机选择删除点。

2.4 算法流程

(1) 种群初始化。本算法中初始种群采用随机方式产生, 路径包含路点数在 $[2, N_{max}]$ 上服从均匀分布, 路径上各路点坐标值在整个地图坐标范围内随机取值, 服从均匀分布。种群规模记为 G_{pn} 。

(2) 采用锦标赛选择法挑选父个体, 并按一定的概率分布随机选择不同的遗传算子以执行遗传操作。

(3) 获得子代个体后检查子代个体染色体长度是否超过 N_{max} , 如果没有超过则保留, 否则丢弃。

(4) 重复 (2)、(3) 步直到获得的新个体数量与父代群体数量相等。

(5) 父代群体和子代群体构成规模为 $2 * G_{pn}$ 的并集, 通过代价函数给出该合集中每个个体的代价值, 选择代价值较低的前 G_{pn} 个个体构成新的父代群体, 取代原有父代群体。

(6) 检查算法停止条件。本文中, 当进化代数 E_{vn} 达到 50 时, 算法结束并选择当前父代群体中的代价值最小个体作为输出。否则跳转至第 (2) 步, 算法继续进行。

3 仿真结果

本文通过多种地图情况下的计算机仿真来验证所提改进算法的性能。在仿真中, 算法相关参数设定如下: 种群规模 100, 进化代数 50, a_1, a_2 分别为 1, 1.000。

3.1 地图 1

地图中, 机器人起始点为 (0, 0), 目标点 (50, 10), 算法运行

多次, 典型结果如图 2, 其中的折线轨迹即为算法的输出路径, 其代价函数值为 109.956 1, 路径全长 109.956 1。

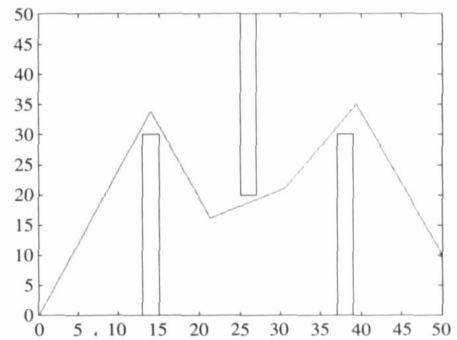


图 2 算法输出结果 1

3.2 地图 2

本实验中机器人起始点 (0, 0), 目标点 (40, 30), 算法运行多次, 典型结果如图 3, 其代价函数值为 80.083 5, 路径全长 80.083 5。

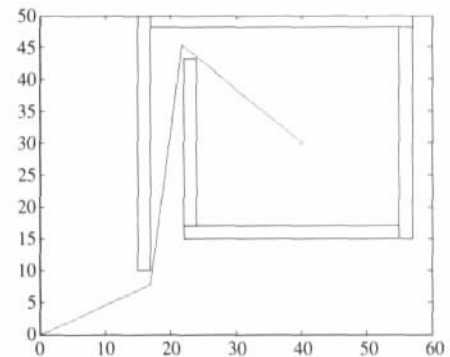


图 3 算法输出结果 2

3.3 地图 3

本实验中机器人起始点 (0, 0), 目标点 (42, 60), 算法运行多次, 典型结果如图 4, 其代价函数值为 76.141 2, 路径全长 76.141 2。

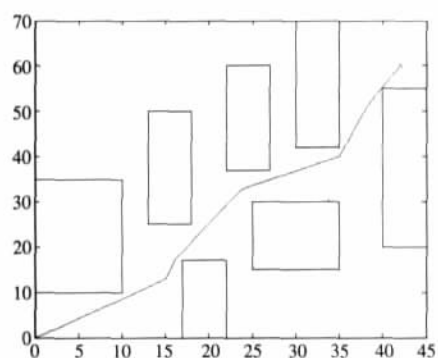


图 4 算法输出结果 3

3.4 算法对比

本实验通过使用不同遗传算子, 比较由此导致的算法性能差异。对比实验中, 算法分为 3 种, 第 1 种为标准遗传算法, 第 2 种使用改进型遗传算子, 但不包含本文中所提的自定义算子, 第 3 种同时使用了本文中提出的改进型遗传算子和自定义遗传算子。其它算法设置完全相同。针对上述 3 幅不同的地图, 每种算法运行 100 次, 当输出结果和前 3 个实验中对应的典型结果的代价值相差 50% 以内时, 认为算法成功。记录算法 100 次运行中的成功率、平均代价值、代价值的标准差, 作为评价算

法优劣的标准。统计结果如表 1~表 3 所示。

	地图 1	地图 2	地图 3
算法 1	57	41	0
算法 2	72	59	0
算法 3	100	99	92

	地图 1	地图 2	地图 3
算法 1	113.234 2	82.580 9	NA
算法 2	111.324 2	81.128 3	NA
算法 3	109.617 8	82.647 8	78.248 5

	地图 1	地图 2	地图 3
算法 1	4.147 1	1.449 4	NA
算法 2	7.560 3	1.454 8	NA
算法 3	5.672 1	3.183 7	10.681 9

4 总结

上述的实验数据证明了本文所提出的改进型遗传算法的有效性。在 3 幅不同的地图上都达到了 90% 以上的算法成功率,并且相对其它算法有明显提高。随着地图的不同,各算法的成功率均出现不同程度波动,但改进型遗传算法波动幅度最小,保持了较好的稳定性,体现出良好的地图适应能力。

本文通过直接基于坐标值的染色体编码方案和有效的代价函数构造使得路径规划中的地图信息被成功引入到遗传操作的实现过程,并由此提高了算法的成功率。值得指出,参数 N_{max} 采用了手动设定的方式来确定。由前面的算法描述可知,过大的 N_{max} 会加大所需的计算量,从而导致输出结果耗时的延长,而过小的 N_{max} 又使算法无法充分的搜索解空间,影响解的优化质量,甚至无法输出有效解。显然,有效的参数优化,将会进一步提高改进型遗传算法的性能。与此类似的还有多点交叉的点数,不同遗传操作的选中概率,染色体变异概率,代价函数

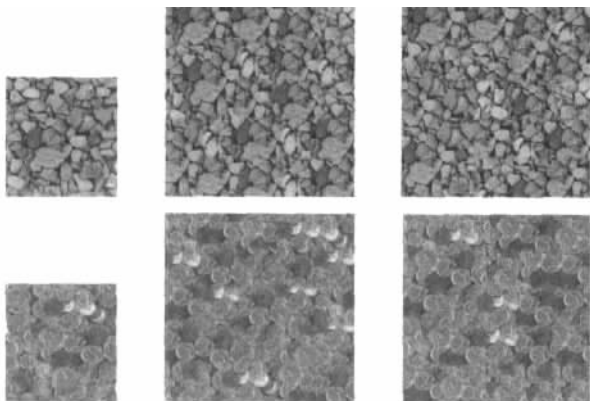
中的权值系数。如何让算法自适应的调整并最优化这些参数的取值,将是本算法的进一步改进方向。

(收稿日期 2006 年 12 月)

参考文献:

- [1] Ayala-Ramirez V, Perez-Garcia A, Montecillo-Puente F J, et al. Path planning using genetic algorithms for mini-robotic tasks[C]// Proceedings IEEE SMC 2004 Conference, Hague, Netherlands, Oct 2004, 4: 3746-3750.
- [2] Tu J, Yang S. Genetic algorithm based path planning for a mobile robot[C]// Proc of IEEE Intl Conf on Robotics and Automation, Taiwan, September 2003: 1221-1226.
- [3] Qi Yuan-qing, Sun De-bao, Li Ning, et al. Path planning for mobile robot using the particle swarm optimization with mutation operator [C]// Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai, August 2004: 2473-2478.
- [4] Hocaoglu C, Sanderson C. Planning multiple paths with evolutionary speciation[J]. IEEE Trans on Evolutionary Computation, 2001, 5(3): 169-191.
- [5] Farritor S, Dubowsky SA. A genetic planning method and its application to planetary exploration[J]. ASME Journal of Dynamic Systems, Measurement and Control, 2002, 124(4): 698-701.
- [6] Wang Chun-miao, Soh Y C, Wang Han, et al. A hierarchical genetic algorithm for path planning in a static environment with obstacles [C]// IEEE Canadian Conference on Electrical and Computer Engineering, Canada, 2002, 3: 1652-1657.
- [7] Hu Yan-rong, Yang S X. A knowledge based genetic algorithm for path planning of a mobile robot[C]// Proceedings of the 2004 IEEE International Conference on Robotics Automation, New Orleans, April 2004: 4350-4355.
- [8] Gerke M. Genetic path planning for mobile robots[C]// Proc of the 1999 American Control Conference, San Diego, CA: IEEE Press, 1999, 4: 2424-2429.

(上接 33 页)



(a) 输入的 128x128 的样本纹理 (b) 运用 PSO 算法输出的 200x200 的目标纹理 (c) 运用 QPSO 算法输出的 200x200 的目标纹理

图 5 两种算法的合成结果比较

搜索过程中具有较大的优势,其收敛速度是标准 PSO 算法的数倍。这是因为 QPSO 所基于的量子系统具有非线性性和不确定性,所以在这样一个系统中,一个粒子能够以某一确定的概率出现在整个可行的搜索空间中任意一个位置,而非限定在某一个固定区域内,因而增加了找到更好适应值的可能。

将 QPSO 引入了纹理合成算法中,与先前的 PSO 纹理合成

比较起来,有收敛速度快,控制参数减少,合成结果的质量更高等优点。但其仍具有许多缺点,如仍需要人为的设定块的大小,迭代次数,以及粒子的数目,对于某些特定的纹理实现效果不佳,以及块算法自身的缺陷。

在未来的工作中,可以将以自适应算法为方向发展,结合 QPSO 算法,实现更好的纹理合成果。

(收稿日期 2007 年 3 月)

参考文献:

- [1] Guo Bai-ning, Liang Lin, Liu Ce, et al. Real-time texture synthesis by patch-based sampling[J]. ACM Transaction on Graphics, 2001, 20(3).
- [2] Shum H Y, Szeliski R. Panoramic image mosaics[R]. Microsoft Research.
- [3] Kennedy J, Eberhart R C. Particle swarm optimization[C]// Proc IEEE Int'l Conf on Neural Networks Vol IV. IEEE Service Center, 1995.
- [4] Sun J, Feng B, Xu W. A global search strategy of quantum behaved particle swarm optimization[C]// IEEE Proc of Conference on Cybernetics and intelligent system, 2004.
- [5] Zhang Yan, Meng Yu, Li Wen-hui, et al. A fast algorithm for image analogy using particle swarm optimization [C]// Proc of the 3rd International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004.

基于改进交叉覆盖设计算法的鲁棒性

宋茂斌¹, 宫宁生^{1,2}

SONG Mao-bin¹, GONG Ning-sheng^{1,2}

1. 南京工业大学 信息科学与工程学院, 南京 210009

2. 南京航空航天大学 信息科学与工程学院, 南京 210016

1. Institute of Information Science and Engineering, Nanjing University of Technology, Nanjing 210009, China

2. Institute of Information Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

E-mail: songmaobin@163.com

SONG Mao-bin, GONG Ning-sheng. Robustness based on ameliorating alternative covering design algorithm. Computer Engineering and Applications, 2007, 43(22): 71-72.

Abstract: This paper gives a method of amelioration to the robust of alternative covering design algorithm based on the analysis of the algorithm. Some instances show that amelioration can obviously enhance the ability of the neural network.

Key words: multi-layer neural networks; alternative covering; robustness

摘要: 在对交叉覆盖设计算法进行分析的基础上, 针对该算法所显示的鲁棒性不强的问题, 给出了一种改进的方法。模拟实验表明了这种改进能够明显地提高神经网络的泛化能力。

关键词: 多层神经网络; 交叉覆盖; 鲁棒性

文章编号: 1002-8331(2007)22-0071-02 文献标识码: A 中图分类号: TP18

1 引言

基于覆盖的构造型前馈神经网络是一种新型神经网络模型^[1], 其基本思想是: 假定训练样本或待识别样本是 n 维向量, 构造型神经网络将所有的训练样本投影变换到 $n+1$ 维的超球面上。这样, 模式识别问题被转化为在 $n+1$ 维的超球面上的覆盖问题。通过在超球面上寻找可以对训练样本集进行正确分类的较优 (即尽可能大的) 覆盖来实现对样本空间的划分, 并根据得到的覆盖来判断待识别样本的类别。该模型具有明显的几何意义, 有可理解性强、识别率高、计算速度快、可以保证训练样本集上 100% 的识别率等优点, 适合于处理大规模的分类任务。在这一算法基础上有许多改进和应用, 有较好的效果^[4,5]。

但是在此模型上设计的交叉覆盖算法往往为尽可能提高学习速度, 而不断寻找最大覆盖, 导致了算法的泛化能力下降。由于实际应用中多注重算法的泛化能力, 因此, 本文对此进行了探讨, 给出了一种提高泛化能力的方法。

2 覆盖算法简介

M-P 神经元模型的几何意义是: 若限定输入向量是在 $n+1$ 维空间的某个球面 S^1 上 (其中心在原点, 半径为 R), 则 $(W^*x - \theta) > 0$ (其中 W 是权向量, θ 是阈值) 就表示球面上落在由超平面 P (其方程为: $(W^*x - \theta) = 0$) 所分割的正半空间的部分, 这个部分恰好是球面上的某个“球形领域”。若取 W 与 x 等长, 则这个“球形领域”的中心恰好是 W , 如图 1 所示。

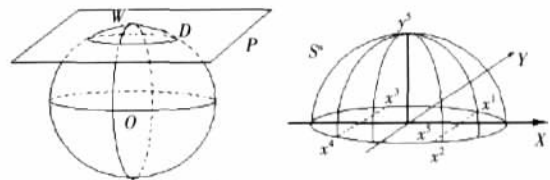


图 1 超平面 P 与超球面相交, 形成“球形区域”的示意图

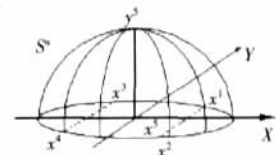


图 2 从 $D \rightarrow S^1$ 变换的示意图

交叉覆盖算法根据上述 M-P 神经元模型的几何意义, 将 n 维空间的样本投影到 $n+1$ 维空间的超球面上, 如图 2 所示。变换法则如下:

$$T: D \rightarrow S^1, x \in D, T(x) = (x, \sqrt{d^2 - |x|^2})$$

其中 $d = \max\{|x| | x \in D\}$ 。

这种变换显然是——对应的。当给定的输入向量的长度不相等时, 可用上面给出的方法, 将它变换成长度相等的情况。

3 覆盖网络的误识分析与改进

根据文献[2], 设样本集 K 为 $K = \{(x_i, y_i) | i = 0, 1, 2, \dots, s\}$, 那么, 所谓“记忆”, 就是当输入为 x_i 时, 网络的输出为 y_i , 记住 x_i 与 y_i 之间的对应关系; 至于“联想”, 就是当输入为 $x_i + \Delta i$ 时, 其输出仍为 y_i (其中 Δi 是噪声或扰动)。可知, 如果对较大的 Δi , 网络仍然能正确输出, 则该网络的“记忆”及“联想”能力强, 即有较强的泛化能力。另一方面, 若将“当输入为 $x_i + \Delta i$ 时, 其输出仍为 y_i ”理解为: 输入落在 x_i 的附近时, 其输出应为 y_i , 否则

其输出就不为 y_i ,若将在 ' x_i 附近'看作是 x_i 的一个球形领域 ,那么 ,由神经元的几何意义知 :若取以 x_i 为中心 ,以 Δ_i 为半径的 '球形领域' ,则其对应的神经元 ,就能完成上述的联想记忆的任务。而 Δ_i 的大小决定了网络泛化能力的强弱。

在求覆盖领域的时候 ,如果使 Δ_i 足够大 ,即保证每一个训练样本都能代表一定的邻域空间 ,就能够使得网络具有良好的泛化能力。本类的覆盖的半径与非本类的样本点之间应该保留至少是样本点的吸引半径的一段距离。

根据文献[1] ,该算法对学习样本的识别率为 100% ,而在测试样本有一定随机扰动时 ,识别率则下降很多。

泛化能力不理想 (抗随机干扰能力不强)的主要原因 是由于每个覆盖的边界接近另外一类 ,即 :在求以 a 为中心的领域 $C(a)$ 时 (详见文献[1]) , $C(a)$ 对应的权和阈值 ($W = (w_j^i) \theta = (\theta_i)$) 是按下面的公式求得 :

$$d^1(i) = \max_{x \in K^1} \{ |a - x| \}$$

$$d^2(i) = \min_{x \in K^2} \{ |a - x| \}$$

$$d(i) = \frac{d^1(i) + d^2(i)}{2} \dots (*)$$

$$\theta_i = d(i)$$

$$W = (a) \theta = (\theta_i)$$

由上述公式知 ,当 $d^1(i) \rightarrow d^2(i)$ 时 $d(i) \rightarrow d^2(i)$,因而发生如图 3 所示的情况 ,导致领域 $C(a)$ 极其接近 K^2 中的点 ,即存在 K^2 中的点 ,该点的吸引域与领域 $C(a)$ 的交集非空 ,因而在测试样本时造成误判 (为直观计 ,将公式中的内积表示成了距离 ,实际含义是相同的) ,而实际上该种情况并非不可区分。

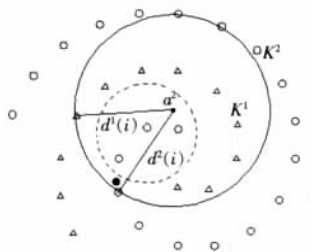


图 3 样本误识分析

其中 ,表示 K^1 类的样本点 ,表示 K^2 类的样本点 ,表示 K^2 类的测试样本误判为 K^1 类的情况。如果考虑扰动样本点 ,容易看出 ,识别率将大大降低。对照图可知 ,正是上述情况导致了误差的出现。

定理 1 设任意一个覆盖领域 $C(a)$ 的中心为 $a(a \in K^1)$,半径为 R ,与其相邻的最接近的 K^2 类的点是 a ,则点 a 的吸引域侵入覆盖领域 $C(a)$ 部分的 '侵入距离' $r < R/2$ 。

证明 覆盖领域 $C(a)$ 的中心是一个 K^1 类点 a ,则 $r < d(a, K^2) = (R+r) / 2$

其中 r 是点 a 到覆盖领域 $C(a)$ 的边界的距离。

而

$$r + r = (R+r) / 2$$

所以

$$r = (R-r) / 2 < R/2 \quad \text{证毕。}$$

实际上 ,由于覆盖的半径越来越大 ,即 R 具有单调增加的趋势 ,而 $r+r$ 固定不变 ,因此 r 相对 $R/2$ 而言是很小的量。因此 ,可能的出错域是一个外径为 R ,内径为 $R-r$ 的一个超圆环域。所以 ,假如将得到的覆盖的半径 R 缩小为 $R-r$,则出错的超圆环域就被避开 ,网络的泛化能力应该能够得到较大的提高。

定理 2 设任一覆盖的样本点与非本类的样本点的最小距离为 D ,则该覆盖为避开超圆环域而回退的距离 $d < D/2$ 。

证明 显然 ,样本点的吸引域 $r_a < D/2$,又由定理 1 知 $r < r_a$,所以 $d < D/2$ 证毕。

显然 ,若 $d < D/2$,则得到的泛化能力与付出的执行效率都成下降趋势 ,属于不予考虑的范围。一般情况下 ,算法的执行效率和泛化能力是矛盾的。随着实际应用中网络的执行效率和泛化能力的要求不同 ,适当调节参数值 d 的大小 ,就可以做到突出网络的执行效率或是其泛化能力。而在实验中 ,通过调整 d 的大小也证明了网络的执行效率和泛化能力之间的这种变化关系。

4 模拟结果

在模拟实验中学习样本的 D 即为螺旋线的径向距离 ,双螺旋线 $D = \dots$,当取 $d = \dots / 2$ 时 ,网络的泛化能力最好 ;而三螺旋线 $D = 2 \dots / 3$,当取 $d = \dots / 3$ 时 ,网络的泛化能力最好。

为了验证上述神经网络改进算法的效果 ,对平面双螺旋线和平面三螺旋线的识别情况进行模拟实验。学习完成后针对螺旋线径向上 20% - 100% 的扰动 ,取样本点进行测试 ,其中 100% 的扰动是指扰动达到最大 ,两类样本点的扰动区域已经相接的状态。每级扰动各进行 1 000 次实验 ,实验结果如表 1、表 2 所示 ,表中列出实验中的平均识别率 ;每次实验每条线取 10 000 个带干扰的测试样本点。另外 ,改进算法对不加扰动的测试样本点的识别率都是 100%。

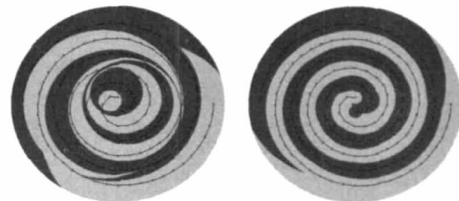
表 1 平面双螺旋线加噪测试情况 (学习样本数 78 × 2)

扰动范围	识别率/%				
	20%	40%	60%	80%	100%
原网络 (10 个覆盖)	95.098 5	92.007 3	88.421 7	84.349 9	80.058 9
改进网络 (21 个覆盖)	99.933 7	99.731 4	99.512 0	99.260 8	96.726 3

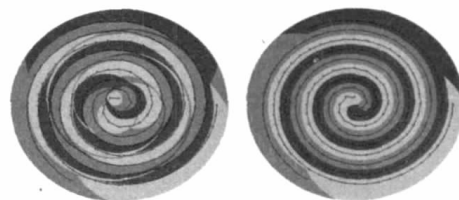
表 2 平面三螺旋线加噪测试情况 (学习样本数 78 × 3)

扰动范围	识别率/%				
	20%	40%	60%	80%	100%
原网络 (26 个覆盖)	89.221 3	81.528 6	76.460 7	72.709 6	69.452 8
改进网络 (35 个覆盖)	99.999 4	99.955 8	99.848 3	99.713 5	97.292 2

改进前后覆盖网络的二维投影如图 4、图 5 所示 ,从图可以看出 ,改进后的网络更好地区分了螺旋线。



(a)改进前 (b)改进后
图 4 平面双螺旋线分类结果



(a)改进前 (b)改进后
图 5 平面三螺旋线分类结果 (下转 108 页)

一种新的基于统计的背景减除方法

孙吉花, 刘肖琳

SUN Ji-hua, LIU Xiao-lin

国防科技大学 机电工程与自动化学院 自动化所 333 教研室, 长沙 410073

333 Staff Room of Automation Institution, College of Mechatronics Engineering and Automation, National University of Defense Technology, Changsha 410073, China

E-mail: sunjh81@yahoo.com

SUN Ji-hua, LIU Xiao-lin. New background subtraction algorithm based on statistic. Computer Engineering and Applications, 2007, 43(22): 73-75.

Abstract: A new effective background subtraction algorithm for color video is proposed. First gets an image sample sequence from a color video. Then analyzes and classifies the RGB value per pixel of the sequence and uses the average RGB value of the class which has the highest frequency in the sequence to construct the background model. At last subtracts background according to both color distortion and brightness range. At the same time, does some morphological work to improve the result. Experiment results indicate that this algorithm can solve the problem that gray video background subtraction algorithms can't detect the targets which have the similar gray intensity compared to background. Furthermore, it can construct background model faster than other color video background subtraction algorithms and allows presence of moving objects in the sample images. It improves the accuracy and real-time performance of background subtraction to a certain extent. Finally proposes some effective methods to update background model so as to adapt the illumination and background changes.

Key words: background model; background subtraction; background update

摘要: 提出了一种有效的彩色视频背景减除的新方法。首先对彩色视频采样得到图像序列, 统计序列中各像素的 RGB 值并归类, 用出现概率最高类的 RGB 均值来构建背景模型, 然后根据颜色差异和亮度范围综合条件, 结合形态学处理进行背景减除。实验结果表明, 此法可以很好地克服灰度视频背景减除中很难识别与背景灰度相近目标的缺陷, 同时比传统的彩色视频背景建模快速, 且样本中允许运动目标存在。对背景减除的准确性和实时性有一定程度地改进。该文还针对光照和背景变化提出了一些有效的背景更新策略。

关键词: 背景建模; 背景减除; 背景更新

文章编号: 1002-8331(2007)22-0073-03 文献标识码: A 中图分类号: TP391

1 引言

视频图像中往往包含大量的信息, 要存储和计算如此庞大的数据集合是很困难的事情。况且真正有用的信息只是其中的很少部分。减除无用的背景就可以大大减少数据量而不丢失有用信息。

背景减除包括背景建模、检测前景和背景更新三个方面。背景建模即从一系列视频图像中准确找到属于背景的部分, 并且存储为背景图像。背景建模面临的困难主要是如何适应光照变化、摄像机抖动、背景中某些物体的高频运动(如树叶舞动和水波荡漾)、背景自身的变化(如本来属于背景的静止对象开始运动, 甚至离开视线或运动对象停止在背景中不动等)^[1]。检测前景是将当前视频图像与背景模型进行比较, 找到前景目标, 减除背景。这里的困难是差值定义和阈值设定的问题, 即反映当前视频与背景差异的度量标准是什么, 分类的阈值为多少。背景更新就是某些变化导致原来的背景模型不再适合当前视

频图像时, 实时地更新背景模型。背景更新的困难在于如何找到合适的更新策略, 能够快速而真实的反映变化, 即实时性和准确性的问题。

本文第 2 章用基于像素 RGB 值统计归类重建背景模型的方法构建初始的背景模型; 第 3 章以颜色差异和亮度范围为依据, 结合形态学处理进行背景减除; 第 4 章提出两种背景更新的方法应对光照变化和物体增减导致的背景变化; 第 5 章给出实验结果和讨论。

2 背景建模

2.1 基本原理

固定摄像机拍摄的视频序列, 背景往往是最经常被看到的。基于这一假设, 参考灰度统计归类方法^[2]对像素 RGB 值统计归类来建立背景模型。即将样本序列中的各个像素的 RGB 值分别进行统计归类, 出现频率最高的像素 RGB 值分别取其

作者简介: 孙吉花, 女, 硕士研究生, 研究方向是模式识别与智能系统; 刘肖琳, 女, 教授, 导师, 主要从事虚拟现实与模式识别方面的研究工作。

平均值作为该像素的背景 RGB 值。

2.2 算法

步骤 1 从一段视频中每隔采样 N 帧图像作为背景建模的样本,记为 $\{A_1, A_2, A_3, \dots, A_N\}$ 。

步骤 2 在 (x, y) 位置分别对各个样本图像的 R 值从 1 到 N 进行统计分析。把 R 值暂且当成灰度值求取灰度直方图,找到直方图的峰值,求峰值为中心的、长度为 e (归类阈值,一般在 10~25 之间)的区间内的灰度平均值,将其作为背景模型中 (x, y) 像素的 R 值。

步骤 3 改变 (x, y) 转到步骤 2。循环遍历所有像素,完成 R 背景建模。

步骤 4 对样本图像的 G, B 值分别进行步骤 2 到步骤 3 的处理,完成 G, B 的背景模型。组合 RGB 值,获得背景模型。

该算法建立背景模型时采样的视频图像中可以包含运动的物体,而且算法的可调参数只有归类阈值 e 一个,应用很简单,不用为多个参数的正确组合而大费周章,省去很多工作量。

3 背景减除

3.1 基本原理

读取当前视频图像各个像素 RGB 值,与背景模型中该像素的 RGB 值进行比较分析。计算颜色差异和判断亮度是否在可接受范围^[3]。如果不同时满足颜色差异小于阈值,而且亮度在可接受范围的综合条件,则该像素为前景目标,直接输出原值。否则,该像素为背景,将其 RGB 值赋 0。

本文计算颜色差异和亮度阈值的颜色模型示意图如图 1。

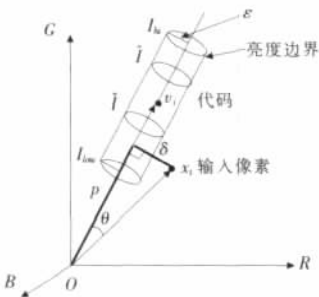


图 1 颜色模型示意图

当前输入像素 $x_i = (R, G, B)$ 的,同一位置的背景像素 $v_m = (\bar{R}_m, \bar{G}_m, \bar{B}_m)$ 根据向量分析的相关知识,计算颜色差异步骤如下:

向量 x_i 的模的平方: $x_i^2 = R^2 + G^2 + B^2$

向量 v_m 的模的平方: $v_m^2 = \bar{R}_m^2 + \bar{G}_m^2 + \bar{B}_m^2$

向量 x_i 和 v_m 的内积的平方: $\langle x_i, v_m \rangle^2 = (\bar{R}_m R + \bar{G}_m G + \bar{B}_m B)^2$

向量 x_i 在 v_m 上的投影的平方 $p^2 = x_i^2 \cos^2 \theta = \frac{\langle x_i, v_m \rangle^2}{v_m^2}$

颜色差异: $= \sqrt{x_i^2 - p^2}$

亮度 I 即向量 x_i 的模: $I = \sqrt{R^2 + G^2 + B^2}$

迭代计算时: $\hat{I} = \max(I, \hat{I}), \check{I} = \min(I, \check{I})$,

亮度范围的上、下界限: $I_{low} = \alpha \hat{I}, I_{hi} = \min\{\beta \hat{I}, \frac{\check{I}}{\alpha}\}$

其中 \hat{I}, \check{I} 表示背景模型中该像素的最小和最大的亮度。 α 关乎亮度范围界限,较小的 α 允许大的亮度范围,较大的 α 允许的亮度范围很窄,一般的取值在 0.4~0.7 之间。 β 的应用是因为大多数情况下阴影都被观测到。 β 的取值在 1.1~1.5 之间。合

适的 α 和 β 取值对背景模型的建立有很大影响。

3.2 算法

在每个像素 (x, y) 位置执行以下步骤:

(1) $x = (R, G, B), I = \sqrt{R^2 + G^2 + B^2}$

(2) 计算 x 与背景中该像素点的 RGB 值之间的颜色差异。比较 I 和 e_i (前景检测阈值,一般在 0.1~1 之间,太大会使训练样本中运动的前景也包含在背景模型中,太小又会影响到运算速度和背景模型本身的性质)。比较 I 与背景中该像素的 I_{low} 和 I_{hi} 的大小。

(3) 如果没有同时满足 e_i 和 $I_{low} < I < I_{hi}$ 两个条件,则该像素为前景目标,此处保留图像本身的 RGB 值。否则即为背景像素,将其值赋 0。

当所有像素都执行这些操作之后,可以得到减除背景之后的前景目标图像。但是这样计算的结果中会含有一些噪声,表现为零星的点块。进行形态学的处理可以较好的解决这一问题。对 RGB 三个分量分别以 3×3 的正方形结构元素进行开运算处理,然后再将三个分量综合显示,可以得到几乎不存在零星点块的视频图像。也可以用滤波降噪的方法来消除这些零星的点块。

4 背景更新

光照变化会引起背景变化,若不及时更新背景,很可能出现过检测、误警,或者降低对前景目标的灵敏度。提高背景模型对全局光照变化的适应性的一种方法是背景更新时加上背景 RGB 向量的平滑处理。用参数 γ (学习率,在 0~1 之间)调整背景 RGB 值,即 $v_m = \gamma * x_i + (1 - \gamma) * v_m$ 。调整 γ 值来适应不同强度的全局光照变化。光照变化大的情况下 γ 值相对大些,变化小的 γ 值相对较小。但这只是缓和的处理,可能很难有理想的效果,要及时建立新的光照条件下的背景模型才是根本的解决方法。判断全局光照变化的方法是统计背景减除过程中前景像素所占的比例,如果大于 80%,则可以看作全局光照变化。连续出现 10 帧大于 80% 的情况,进行背景更新。更新的方法是重新采样视频图像,按先前的背景建模方法建立新的背景模型。

物体停止在检测范围变成背景,或者背景中有物体开始运动变成前景,都会导致背景发生局部的变化。若对整个视野范围进行背景更新,需要很大的计算复杂度,而且变化的区域往往不会是背景的大部分,不需要整个背景模型的更新。只需要找到变化的区域进行更新,再与原来的背景模型合并以得到新的背景模型。这是基于分层建模的背景更新方法。寻找物体停留的位置,判断物体的大小,对该位置和大小区域用物体成像代替原来的背景。判断物体是否从背景变成前景或从前景变成背景,可以对背景减除后的结果进行比较,如果连续在同一位置出现同一物体 10 次,即有 10 帧图像中存在静止的前景目标,这可以判断为误检。必须进行模型更新了。把静止的物体当作这一位置的背景,简单有效的更新背景模型。

5 实验结果及讨论

在 P4 1.8 G 处理器,256 M 内存, MATLAB7.1 平台上,选取三组视频(大小分别为 $300 \times 200, 334 \times 246, 334 \times 246$),分别采样 30 帧彩色视频图像建立背景模型,对各组视频的 600 帧图像进行背景减除实验,对比分析本文提出基于像素 RGB 统计分析进行背景减除的方法与代码本方法和灰度统计归类方法的优劣。实验结果如图 2。

表 1 实验参数表

参数 个数	采样 帧数	第一组			第二组			第三组			能否存 在运动 物体	能否识别色彩 不同而灰度接 近的目标	
		建模 时间/s	减除频率/ (帧/s)	误检 率/%	建模 时间/s	减除频率/ (帧/s)	误检 率/%	建模 时间/s	减除频率/ (帧/s)	误检 率/%			
本文方法	4	30	9	3.7	12	13	2.7	13	14	2.8	13	能	能
CB方法	4	30	21	3.7	15	33	2.7	15	33	2.8	15	能	能
灰度方法	1	30	4.1	34	不定	5.7	21	不定	5	21	不定	能	不能

(注:减除频率的计算方法是 50 帧视频图像处理的平均速率。误检率的计算方法是 50 帧视频图像处理过程中,统计检测错误的像素占理想的前景目标的像素总数的百分比,取其平均值作为误检率的估计,只是一个约数。灰度统计方法的误检率因为其无法分辨出颜色不同而灰度接近的目标,所以有很大的波动,无法估计具体的数值,取决于是否存在这样的前景目标。)

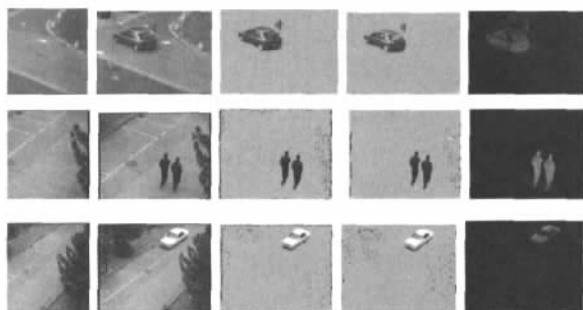


图 2 实验结果图

(从左至右依次为:背景模型、当前图像、本文方法结果、CB方法结果、灰度统计方法结果。)

在视频图像中加入颜色不同而灰度接近的物体,如图 3。



图 3 实验结果图

(从左至右依次为:当前图像、本文方法结果、灰度统计方法结果)

实验过程和结果中的一些参数如表 1。

从实验结果看出,本文提出的方法是很有效的。建模时间少于代码本(CB)方法的一半,而检测目标的准确率高于代码本方法和像素灰度统计方法。实现了时间与准确率的较好结合。

虽然代码本方法可以较好地实现检测目标的准确性,但是建模所需的时间是一个很大的问题,影响了背景更新的速度,导致实时性较差。本文的方法建模时间明显少于代码本方法,而且减除背景的依据是相同的,所以保持了准确性,甚至因为背景模型的优越而提高了准确性。

灰度统计方法很难检测那些颜色差异很大但灰度差异很小的前景目标,本文的方法是彩色视频图像的背景减除,可以

避免这一情况。而前景与背景很少有 RGB 三个分量,甚至亮度都接近的情况,所以漏检的可能性很小。而且比起灰度视频图像背景减除还有一个很大的优点,即可以清楚的看到运动前景的真实色彩,为下一步的监视和跟踪提供了很完整的信息。

背景中有物体停留或离开,的情况下,本文的算法仍存在缺陷。虽然可以进行背景更新,但是变化的背景往往只是整个背景的很小部分,重新建立背景模型耗时太大,况且没有必要。分层建模是一个可以弥补这一缺憾的方法。由浮云等带来的短时全局光照的变化,会导致过检测。抑制这一不利因素的可行方法有对背景像素的 RGB 向量和方差的平滑,或者统计数值的整体偏移处理^[4]。这些都是今后工作要解决的问题。

但是所有统计方法都有一个局限性,那就是建立背景模型时需要采样大量的图像,本文的方法也不例外,这对背景建模和更新的实时性是一个障碍。而且,还难以消除长期缓慢运动的物体对背景建模的影响,使得基于统计的方法的应用受到一定程度的限制。

6 结束语

针对彩色视频的背景减除问题,本文提出了基于 RGB 的统计量建立背景模型和依据颜色差异和亮度范围进行背景减除的方法。有效的解决了灰度统计方法不能识别颜色差异大而灰度接近背景的目标问题,解决了代码本方法建立背景模型速度较慢的问题。实验结果显示,本文的方法提高了彩色视频背景减除的实时性和准确率。(收稿日期 2006 年 12 月)

参考文献:

- [1] Piccardi Massimo. Background subtraction techniques a review[R]. The ARC Centre of Excellence for Autonomous Systems (CAS) Faculty of Engineering, UTS April 15, 2004.
- [2] 侯志强, 韩崇昭. 基于像素灰度归类的背景重构算法[J]. 软件学报, 2005, 16(9): 1568-1576.
- [3] Kim K, Chalidabhongse T H, Harwood D, et al. Real-time foreground-background segmentation using codebook model [EB/OL]. Real-Time Imaging. www.elsevier.com/locate/rti.
- [4] Hong D, Woo W. A background subtraction for a vision-based user interface[C]//ICICS-PCM, Singapore, 15-18 Dec 2003.

131-136.

- [6] 杜琼, 周一. 新的进化算法——文化算法[J]. 计算机科学, 2005, 32(9): 142-144.
- [7] 熊伟清, 魏平, 赵志煜. 遗传算法的一个调节算子的研究[J]. 小型微型计算机系统, 2003, 24(3): 531-533.
- [8] 张铃, 张钊. 遗传算法机理的研究[J]. 软件学报, 2000, 11(7): 945-951.

(上接 40 页)

- [3] 王磊, 潘进, 焦李成. 免疫算法[J]. 电子学报, 2000, 28(7): 74-78.
- [4] 王煦法, 张显俊, 曹先彬, 等. 一种基于免疫原理的遗传算法[J]. 小型微型计算机系统, 1999, 20(2): 117-120.
- [5] Robert R G. An introduction to cultural algorithms[C]//Sebalck A V, Fogel L J, River Edge. Proceeding of the Third Annual Conf on Evolution Programming. NJ: World Scientific Publishing, 1994:

一种新的二值图像连通区域准确标记算法

刘贤喜 李邦明 苏庆堂 刘中合 王玉亮 杨 峰

LIU Xian-xi ,LI Bang-ming ,SU Qing-tang ,LIU Zhong-he ,WANG Yu-liang ,YANG Feng

1.山东农业大学 机械与电子工程学院 ,山东 泰安 271018

2.鲁东大学 ,山东 烟台 264025

1.College of Mechanical & Electrical Engineering ,Shandong Agricultural University ,Tai 'an ,Shandong 271018 ,China

2.Ludong University ,Yantai ,Shandong 264025 ,China

E-mail :bullish@sdau.edu.cn

LIU Xian-xi ,LI Bang-ming ,SU Qing-tang ,et al.New exact labeling algorithm of connected regions in binary images. Computer Engineering and Applications 2007 43 (22) :76- 78.

Abstract : The basic characteristic of current labeling algorithms are analyzed and a new algorithm for exact labeling of connected regions in binary images is put forward.First ,scan and label all the pixel points in turn and judge the labeling conflicts exactly occurring during the labeling source and build the models of them ,moreover add the rescan algorithm in order to eliminate the labeling errors aroused by label conflicts.Experimental result shows that the algorithm can label connected regions of any shape exactly and compared with the current labeling algorithms ,this one has lower rate of repeated scan and higher efficiency and more exact veracity in speed ,so it has wider application.

Key words : binary image ,connected region ,region labeling ,model

摘 要 :在分析已有区域标记算法的基础上 ,提出了一种新的二值图像连通区域准确标记算法。顺序扫描和标记二值图像的各个像素点 ,准确判断标记过程中出现的标记冲突 ,并建立标记冲突的模型 ,在算法中增加回溯扫描算法 ,消除标记冲突引起的标记误差。实验证明该算法可以准确标记出各种形状的连通区域 ,和已有算法相比扫描重复率低、运行准确、速度快 ,具有很好的应用前景。

关键词 :二值图像 ;连通区域 ;区域标记 ;模型

文章编号 :1002- 8331 (2007)22- 0076- 03 文献标识码 :A 中图分类号 :TP391.41

1 引言

经过阈值分割的二值图像 ,各个目标被分割成彼此分离的连通区域 ,对连通区域的标记在机器视觉领域里是一种非常重要的处理方法 ,连通区域标记过程就是要寻找图像中所有的目标对象 ,并且将属于同一目标对象的所有像素用唯一的标记值进行标记^[1]。因此通过某种方法把各个连通区域区分开来 ,分别研究各个连通区域的性质是提取图像特征、进行目标检测和识别的重要一步。

目前 ,连通区域标记法经常采用的方法主要有以下几种 :第一种是边界跟踪—区域填充的方法^[2] ,即先对连通区域进行边界跟踪 ,找到区域的边界并标记 ,然后搜索并标记连通区域的内部点 ,以达到标记不同连通区域的目的。此方法主要的难点在于如何准确实现多轮廓跟踪和填充 ,目前解决的方法主要是采用限制区域的方法 ,此法对目标点距离很近且形状不规则的情况处理不理想 ,准确率低。第二种是区域生长法^[3] ,此法先对图像进行顺序扫描 ,然后每遇到一个目标点 ,就分配一个未使用的标号 ,然后对目标点的邻域进行检测 ,如果有没有标记

的目标点 ,则赋予相同的标号 ,反复进行此操作 ,直到图像扫描结束。这种方法重复扫描次数较多 ,对连通区域形状的适应性不强。第三种是跟踪算法^[4] ,二值图像中每个目标点都被标记为一个与其坐标相关的标号 ,然后扫描标记后的图像 ,并将每个像素的标号改为其邻域内的最小标号。反复执行此操作 ,直到不需要作标记更改为止。此方法处理小而凸的目标时 ,收敛速度较慢 ,处理效果差。

本文在认真分析以上算法的基础上 ,提出了一种顺序扫描二值图像进行标记的算法。首先分析了在扫描标记过程中由于连通区域的形状不规则引起的标记冲突 ,并建立了标记冲突模型 ,然后在标记算法中增加回溯扫描算法对选择的两种典型情况的标记冲突进行处理 ,并对其它情况的标记冲突进行了分析说明 ,实验证明算法可以准确标记出各种形状的连通区域。

2 顺序扫描标记算法中的标记冲突

设经过阈值分割的二值图像的背景点的像素值为 BJ ,目标点的像素值为 MB ,扫描图像的方向为从左到右 ,从下到上 ,

基金项目 :山东省教育厅资助科研课题 (the Research Project of Department of Education of Shandong Province ,China under Grant No.J05B07)。

作者简介 :刘贤喜 (1963-) ,男 ,博士 ,教授 ,硕士生导师 ,研究方向 :图像处理、计算机视觉和模式识别技术 ;李邦明 ,硕士研究生 ;苏庆堂 ,讲师 ;刘中合 ,硕士研究生 ;王玉亮 ,硕士研究生 ;杨峰 ,硕士研究生。

即从图像的左下端开始。dq代表当前扫描点(坐标为(i,j))和它的像素值。lt、tt、rt、ll、rr、lb、bb、rb分别代表与dq相连通的8邻域的像素左上、上、右上、左、右、左下、下、右下点及其像素值,具体各点的位置与坐标如表1所示。

表1 8邻域各点位置及坐标

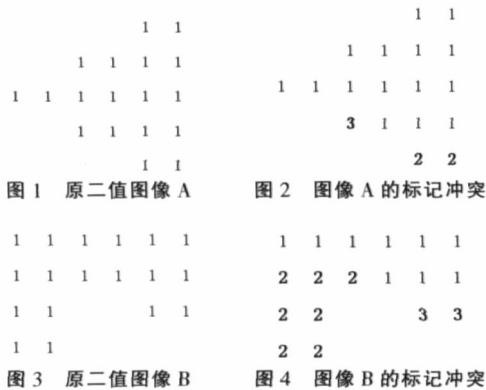
lt (i+1, j-1)	tt (i+1, j)	rt (i+1, j+1)
ll (i, j-1)	dq (i, j)	rr (i, j+1)
lb (i-1, j-1)	bb (i-1, j)	rb (i-1, j+1)

顺序扫描标记法的思路是:从图像左下端开始扫描,如果dq点的像素值为BJ,继续扫描;如果点dq的像素值为MB,则检查已经扫描过的ll、lb、bb、rb四点的像素值,设tag为像素点的标记初值(初始值设为2),作以下两种处理:

(1)如果ll、lb、bb、rb的像素值皆为背景色BJ,即扫描过的周围点没有被标记过,认为当前点dq为连通区域的第一个被扫描的点,则将标记为tag,即dq=tag,然后tag自动加1;

(2)如果ll、lb、bb、rb的像素值不全是背景色BJ,即ll、lb、bb、rb部分(或全部)被标记,且标记值相同,设为tag,认为当前点和周围点属于同一个连通区域,则将dq点标记为此标记值,即dq=tag。

上述处理是为了保证相邻像素点的标记值相同,但在扫描的过程中,会出现ll、lb、bb、rb不同标记值的情况。如图1和图3所示,原图像A和B本来分别只有一个连通区域,但经顺序扫描标记后会出现如图2和图4所示的标记冲突,其中BJ=0(省略),MB=1、2、3为标记值,导致同一个连通区域内出现不同的标记值,即出现标记冲突,它破坏了连通区域的连通性,也为后续处理带来了困难,为实现区分连通区域的目的,标记冲突必须消除。



3 回溯扫描算法

图2和图4标记冲突的出现,都是由于当dq=MB并且ll=lb=bb=rb=BJ时,在同一个连通区域产生新的标记值而形成的,要直接从根源上消除标记冲突,比较困难。间接解决相对较易,当dq=MB并且周围的标记值不同时,即标记冲突出现时,增加回溯扫描算法,把错误的标记值修正为同一个连通区域的最初标记值,可实现连通区域的准确标记。即在上述顺序扫描处理的基础上,增加回溯扫描处理。

如果dq点的ll、lb、bb、rb各点的像素值不全是背景色BJ,即ll、lb、bb、rb都被标记或者部分标记,但标记值不同,设为tag₁、tag₂、tag₃、tag₄,且tag₁≠tag₂≠tag₃≠tag₄,增加回溯扫描算法修改各标记值为tag₁,并将dq点标记为此标记值,即dq=tag₁。

3.1 当ll和rb出现标记冲突时的回溯扫描算法

图2的标记冲突属于ll和rb标记冲突的一种情况,即lb=BJ,bb=BJ,ll=BJ,rb=BJ,且ll>rb,如图5所示为更一般的此种情况的标记冲突的模型,图中BJ=0(省略),MB=1,a,b为标记值,显然b>a;图4的标记冲突属于ll和rb标记冲突的另一种情况,即lb=BJ,bb=BJ,ll=BJ,rb=BJ,且ll<rb,如图6所示为更一般的此种情况的标记冲突的模型,图中BJ=0(省略),MB=1,a,b为标记值,显然b>a。研究图5和图6中的标记冲突,不难发现较大的标记值b是错标的,为了消除标记冲突对后续处理的影响,必须消除大的标记值,即修改大的标记值b为小的标记值a,采用如图7和图8所示的回溯扫描算法可以完成此任务。



图5 ll>rb时标记冲突模型 图6 ll<rb时标记冲突模型

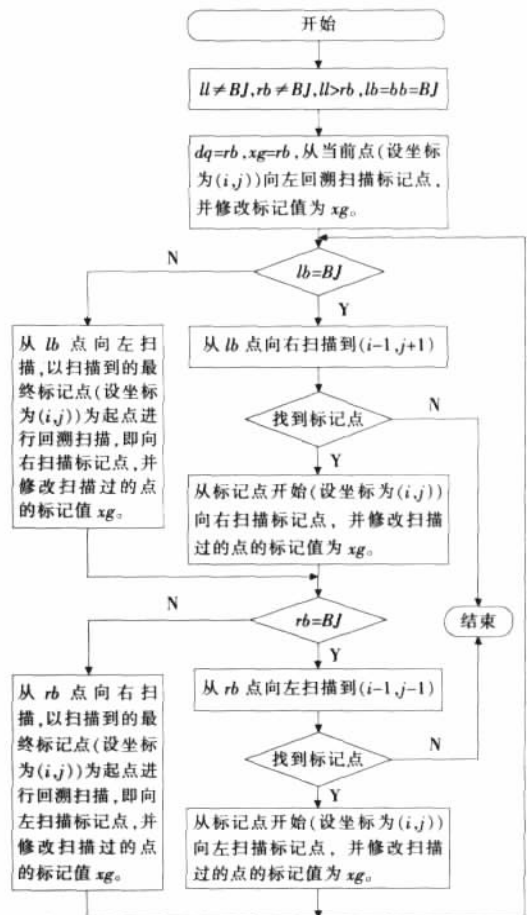


图7 ll>rb时回溯扫描算法流程图

3.2 出现标记冲突其它情况的分析

根据ll、lb、bb、rb标记值之间的相互关系,标记冲突出现的情况很多,仅仅考虑ll、lb、bb、rb点中只有两个点被标记而且标记值冲突的情况,就有C₂⁴种。除去上面介绍的ll和rb标记

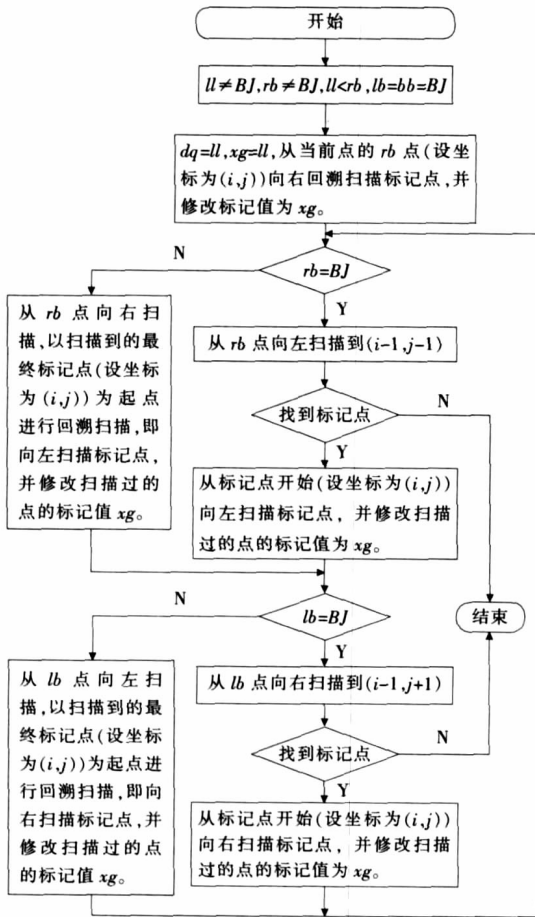


图8 ll < rb 时回溯扫描算法流程图

冲突, 还有 ll 和 lb、ll 和 bb、lb 和 bb、lb 和 rb、bb 和 rb 之间的冲突, 它们的冲突模型分别如图 9-图 13 所示。

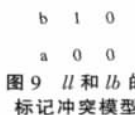


图9 ll 和 lb 的标记冲突模型

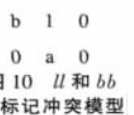


图10 ll 和 bb 的标记冲突模型

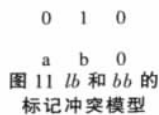


图11 lb 和 bb 的标记冲突模型

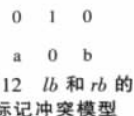


图12 lb 和 rb 的标记冲突模型

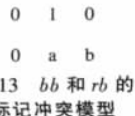


图13 bb 和 rb 的标记冲突模型

虽然标记冲突有以上 5 种类型, 但只要能够保证顺序扫描的第二种处理原则, 即“如果 ll、lb、bb、rb 都被标记或者部分标记, 且标记值相同, 认为当前点和周围点属于同一个连通区域, 则用相同的标记值标记 dq 点”, 从上面的模型可知 ll 和 lb、ll 和 bb、lb 和 bb、bb 和 rb 之间的标记冲突不会出现, 仅有 lb 和 rb 的标记冲突会出现。

当 ll=BJ, bb=BJ, lb=BJ, rb=BJ, 且 lb > rb 时, 按照 ll 和 rb 标记冲突时回溯扫描的处理原则, 需要将连通区域中和 lb 标记值相同的像素点的值改为和 rb 相同的标记值, 这种情况的扫描算法类似于 ll 和 rb 标记冲突的第一种情况, 即 lb=BJ, bb=BJ, ll=BJ, rb=BJ, 且 ll > rb 时的扫描算法, 只不过在 ll 和 rb 标记冲突时, 算法是从发生标记冲突的当前点开始向左扫描, 而在这里, 当 lb 和 rb 发生标记冲突时, 算法需要从发生标记冲突的点的 lb 点开始向左扫描, 其它算法处理完全一样。

当 ll=BJ, bb=BJ, lb=BJ, rb=BJ, 且 lb < rb 时, 按照上面回溯

扫描的原则, 需要将连通区域中和 rb 标记值相同的像素点的值改为和 lb 相同的标记值, 这种情况的回溯扫描算法流程和 ll 和 rb 标记冲突时第二种情况类似, 只不过在 ll 和 rb 标记冲突时, xg=ll, 而当 lb 和 rb 发生标记冲突时, xg=lb。

理论上 ll、lb、bb、rb 之间出现的标记冲突的情况, 还有 3 个或 4 个标记值相互冲突的情况, 但是不管什么类型的 3 个或者 4 个标记值相互冲突的情况都是从两个标记值相互冲突开始的。由于上面对所有两种标记值相互冲突情况的研究和处理, 都是发现标记冲突, 立即进行回溯扫描修正标记值, 消除了标记冲突给后续处理带来的麻烦, 所以只要能够保证按照上述回溯扫描算法处理两种标记值相互冲突的情况, 其它各种情况则不会出现。

4 实验与分析

随着图像采集和处理技术的发展, 利用计算机视觉进行玉米种子的质量检测逐渐受到人们的重视。如何快速准确地提取玉米种子的外形轮廓是进行玉米种子等级评定地一个重要环节。在 VC 环境下, 利用上述标记算法, 对采集到的玉米种子图像进行试验, 结果表明算法可以很好的消除标记冲突, 准确的标出图像中任何形状的各个连通区域 (如图 14 和图 15 所示)。

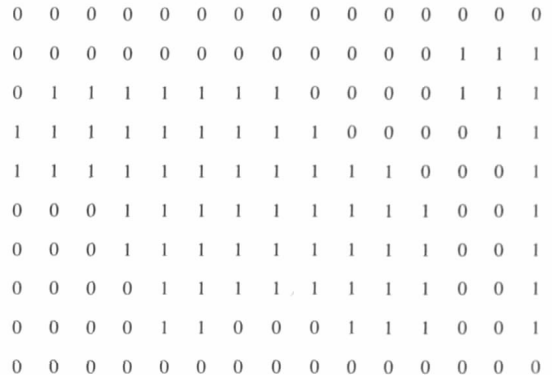


图14 玉米种子部分图像(处理前)

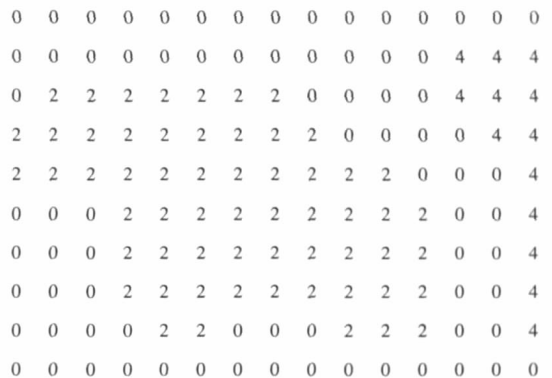


图15 玉米种子部分图像(处理后)

5 结论

利用顺序扫描标记法进行区域标记时, 标记冲突现象经常出现, 本文仔细研究了标记冲突出现的原因, 并通过常见标记冲突的模型对所有的标记冲突进行了分析研究, 提出了回溯扫描算法, 算法在标记冲突出现时, 立即进行回溯扫描, 对引起误差的标记值进行及时修改, 这在一定程度上减少了算法重复扫描的次数, 提高了算法的运行效率, 利用算法对玉米种子图像进行处理试验, 结果表明算法能够准确标记任意复杂形状的连

(下转 98 页)

基于偏好一致性的群体聚类算法

胡立辉

HU Li-hui

长沙理工大学 计算机与通信工程学院,长沙 410076

College of Computer & Communication Engineering, Changsha University of Science & Technology, Changsha 410076, China

E-mail: xhulh@163.com

HU Li-hui. Algorithm for group clustering based on coherence of preference. Computer Engineering and Applications, 2007, 43 (22): 79-82.

Abstract: This paper puts forward to the concept of similarity between member and member set, proves and analyses some useful propositions and corollaries, gives an if and only if condition which can hold the coherence of cluster set when a member joined to the cluster set, and describes a heuristic iterative algorithm for group clustering. The result shows that the algorithm has better clustering performance and lower coherent deviation index.

Key words: group clustering; similarity; coherence; deviation

摘要: 提出了成员与成员集的相似度概念, 证明与分析了一些有用的命题和推论, 给出了一个保持聚集一致性的成员与聚集相聚的充分必要条件, 描述了一个群体聚类的启发式迭代算法。实例测试表明, 提出的算法有较好的聚类性能和较低的一致性偏差指标。

关键词: 群体聚类; 相似度; 一致性; 偏差

文章编号: 1002-8331(2007)22-0079-04 文献标识码: A 中图分类号: TP311; C934

在向量空间的聚类问题中, 常用向量间的夹角余弦或修正的夹角余弦作为向量间的相似度, 并以此作为向量与向量集相聚的判断依据^[1-4]。文献[1, 2]研究了项目协同问题中应用余弦相似度进行聚类的算法, 文献[3]介绍了余弦相似度在图像分类中的应用。文献[4]将修正的夹角余弦相似度应用到群决策的群体聚类问题中, 其算法的思路为: 给出一个相聚判定阈值, 如果候选向量与聚集的代表向量(一般是聚集的平均值中心点)的相似度不低于阈值, 则候选向量可以添加到聚集中。这样计算出的结果, 常常不能保证聚集的一致性指标, 从而降低聚集的整体一致性(或相似性)以及聚类效果。本文在群体偏好向量相似度的基础上, 提出偏好向量与成员集的相似度概念, 得出了候选向量与聚集中心点相聚的修正阈值, 同时也证明了聚集中心点作为聚集的代表或偏好向量的合理性。同时, 本文利用了统计学中的标准差和标准差系数, 分别作为聚集和群体的偏差指标和相对偏差指标, 用以评估群体聚类结果。

本文第1章介绍一些基本的定义, 如相似度、一致性、偏差等; 第2章证明了几个有用的命题和推论, 作为本文算法及相关概念的理论依据; 第3章给出并分析了一个群体聚集的启发式迭代算法; 第4章以实例测试结果, 比较并验证了本文算法的有效性; 第5章对本文的工作进行了总结。

1 术语与定义

本文始终假设, 群体 Ω 的成员由分量非负的 n 维空间中

的向量组成, Ω 非空且不含分量全0的成员, 文中的向量范式为欧氏2-范式并记为 $\|v\|$, 向量 u 与 v 的点积记为 $u \cdot v = \sum_{k=1}^n u_k v_k$ 。此外, 有关偏好向量、相似度等概念在群决策及群体中的解释与意义, 可参见文献[4]。

定义1 n 维向量空间中的偏好向量 u 与 v 的夹角余弦定义为它们的相似度, 记为: $(u, v) = \frac{u \cdot v}{\|u\| \|v\|}$ 。偏好向量 v 与向量集 V 的相似度定义为该向量与集合中所有成员的相似度的平均值, 记为: $(v, V) = \frac{1}{m} \sum_{i=1}^m (v, v^i)$, 其中 m 为集合的成员个数, $v^i \in V, 1 \leq i \leq m$ 。(2)个成员的偏好向量集 $V = \{v^1, v^2, \dots, v^m\}$ 的一致性指标定义为该集合内偏好向量之间的相似度的平均值, 记为: $(V) = \frac{1}{C_m^2} \sum_{1 \leq i < j \leq m} (v^i, v^j)$, 其中向量 $v^i, v^j \in V, C_m^2 = \frac{m(m-1)}{2}$ 。特别地, 如果向量集 V 的成员个数 $m=1$, 则令 $(V) = 0$ 。另外, 还记向量集 V 中的向量与 V 的相似度的平均值为: $\bar{(V)} = \frac{1}{m} \sum_{i=1}^m (v, V)$ 。

定义2 定义 $\bar{V} = \frac{1}{m} \sum_{i=1}^m \frac{v^i}{\|v^i\|}$ 为偏好向量集 V 的几何中

基金项目: 湖南省交通厅项目(No.200610); 长沙理工大学2006年度基金。

作者简介: 胡立辉(1965-)男, 副研究员, 主要从事泛数据库理论与应用研究。

心点,简称中心点。

定义3 偏好矢量集 V 的成员之间的相似性与集合的一致性指标的标准差(也称均方差)定义为该集合的偏差指标,记为:

$$\sigma(V) = \sqrt{\frac{1}{C_m^2} \sum_{i,j=1}^m (v_i, v_j - (V))^2}$$

偏好矢量集的相对偏差指标(也称为标准差系数)定义为: $(V) = \sigma(V) / (V) \times 100\%$ 。

定义4 设 C1, C2, ..., Ck 为群体 Ω 的聚类算法产生的 k 个聚集,则群体 Ω 的一致性指标定义为其聚集的一致性指标的加权平均,记为: $(\Omega) = \frac{1}{m} \sum_{i=1}^k n_i (C_i)$ 其中 ni 为聚集 Ci 的成员个数, m 为 Ω 的成员个数。

聚集的一致性指标与群体的一致性指标的加权标准差定义为群体的偏差指标,记为 $\sigma(\Omega) = \sqrt{\frac{1}{m} \sum_{i=1}^k n_i (C_i - (\Omega))^2}$ 。

群体的相对偏差指标为 $(\Omega) = \sigma(\Omega) / (\Omega) \times 100\%$ 。

上述定义中,聚集及群体一致性指标来源于文献[4],偏差指标、相对偏差指标则取自统计学上的标准差与标准差系数^⑨。由于文献[1,2]只定义了成员之间的相似性,没有全体对象的一致性指标概念,故只给出了绝对偏差的定义。显然,从统计学的观点考虑,相对偏差比绝对偏差更能反映指标的偏离程度。

需要指出,文献[1,2,4]用偏好矢量的修正夹角余弦值作为相似性,但显然存在一个缺陷,即如果矢量的各个分量值全部相同,此时矢量各分量的平均值与分量值相等,相减后将得到一个 0 矢量,不能计算该矢量与其它矢量的相似性。

2 性质与命题

记常数 λ 与矢量 v 的积 λv 为 v 的各个分量乘以常数 λ 后得到的矢量, λV 为集合 V 中的每个矢量乘以 λ 后组成的新矢量集合。

命题1 设 u, v 为 n 维偏好矢量, V 为矢量集, λ = 0 为常数,则如下结论成立:

- (1) $(u, v) = (v, u) = (\lambda u, v) = (u, \lambda v)$;
- (2) $(V) = (\lambda V)$, $(u, V) = (\lambda u, V) = (u, \lambda V)$;
- (3) $(u, v) \in [0, 1]$, $(V) \in [0, 1]$, $(u, V) \in [0, 1]$;
- (4) $(u, v) = 1$ 当且仅当 $u = \lambda v$ 。

证明 结论(1)可以由矢量的相似性定义1直接得到,结论(2)容易由定义1及结论(1)得出,结论(3)则是矢量点积(也称内积)中的经典结果——Cauchy-Schwarz 不等式^⑩,结论(4)可由定义1、结论(1)和 Cauchy-Schwarz 不等式导出。

命题1的结论(1) (2)表明,夹角余弦仅与矢量空间中矢量的方向有关,各个分量同时以常数变化时,并不影响矢量间的相似性及矢量与矢量集的相似性。因此,对群体成员进行规范化(范数值为 1)后,也不影响成员之间的相似性,从而也不会影响以相似性作为相聚条件的群体聚类算法,结论(3)呈现了相似性、一致性指标的[0, 1]取值区间的特性,结论(4)则表明,当且仅当偏好矢量在矢量空间中的方向一致(夹角为 0)时,相似性取最大值。

命题2 给定阈值 $\gamma (0 < \gamma < 1)$,如果矢量集 V 的成员个数 $m=1$ 或者一致性指标 $(V) > \gamma$,且矢量 v 与集 V 的相似性 $(v, V) > \gamma$ 。那么 v 添加到集 V 后,仍然成立 $(V) > \gamma$ 。

证明 当 V 的成员个数 $m=1$ 时,根据定义易知结论成立。

如果 $(V) > \gamma$ 且 $(v, V) > \gamma$,即 $\sum_{i,j=1}^m (v_i, v_j) > C_m^2 \gamma$, $\sum_{i=1}^m (v_i, v) > m\gamma$ 。这样,矢量 v 添加到集 V 后,新集合的一致性指标满足:

$$(V \cup \{v\}) = \frac{1}{C_{m+1}^2} \left(\sum_{i=1}^m (v_i, v_i) + \sum_{i,j=1}^m (v_i, v_j) \right) + \frac{1}{C_{m+1}^2} (m\gamma + C_m^2 \gamma) > \gamma$$

命题2指出,如果矢量集的一致性指标满足阈值 γ ,候选偏好矢量与集合的相似性满足阈值 γ 。那么,添加该矢量到集合后,仍然能保证矢量集的一致性指标满足阈值 γ 。显然,命题2要求计算候选矢量与集合中每个成员的相似性。后文中将证明(推论2),只需要计算候选矢量与中心点的相似性,即可判定其是否满足相聚条件。

命题3 设 $\bar{V} = \frac{1}{m} \sum_{i=1}^m \frac{v_i}{v}$ 为偏好矢量集 V 的中心点, m 为矢量集 V 的个数, $v^j, 1 \leq j \leq m$ 是 V 中的矢量, v 为一偏好矢量,则如下结论成立:

- (1) $(\bar{V}, V) = \frac{1}{m} \sum_{i=1}^m (v_i, V) = \frac{1 + (m-1)(V)}{m}$
- (2) $(\bar{V}, V) = \bar{V} = \sqrt{\frac{1 + (m-1)(V)}{m}}$
- (3) $(v, V) = \sqrt{\frac{1 + (m-1)(V)}{m}} (v, \bar{V})$
- (4) $(v, V) = \frac{2}{m+1} \left(\sqrt{\frac{1 + (m-1)(V)}{m}} (v, \bar{V}) + \frac{m-1}{2} (V) \right)$

证明 先假定矢量集 V 的个数 $m \geq 2$ 。现证明

(1) V 中的矢量与 V 的相似性的平均值为:

$$(\bar{V}) = \frac{1}{m} \sum_{i=1}^m (v_i, V) = \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{m} \sum_{j=1}^m (v_i, v_j) \right) = \frac{1}{m^2} \left(\sum_{i=1}^m (v_i, v_i) + 2 \sum_{i,j=1}^m (v_i, v_j) \right) = \frac{1}{m^2} (m + 2C_m^2 (V)) = \frac{1 + (m-1)(V)}{m}$$

(2) 根据相似性、中心点的定义,有

$$(\bar{V}, V) = \frac{1}{m} \sum_{i=1}^m (\bar{V}, v_i) = \frac{1}{m} \sum_{i=1}^m \sum_{k=1}^n \frac{\bar{V}_k \cdot v_k^i}{\bar{V} \cdot v_i} = \frac{1}{m \bar{V}} \sum_{i=1}^m \sum_{k=1}^n \frac{\bar{V}_k \cdot v_k^i}{v_i} = \frac{1}{m \bar{V}} \sum_{i=1}^m \sum_{k=1}^n \left(\frac{1}{m} \sum_{j=1}^m \frac{v_k^j}{v_j} \right) \frac{v_k^i}{v_i} = \frac{1}{m^2 \bar{V}} \sum_{i=1}^m \sum_{k=1}^n \sum_{j=1}^m \frac{v_k^j v_k^i}{v_i v_j} = \frac{1}{m^2 \bar{V}} \sum_{k=1}^n \sum_{i=1}^m \sum_{j=1}^m \frac{v_k^i v_k^j}{v_i v_j} = \frac{1}{m^2 \bar{V}} \sum_{k=1}^n \left(\sum_{i=1}^m \frac{v_k^i}{v_i} \right)^2 =$$

$$\begin{aligned} & \frac{1}{\bar{V}} \sum_{k=1}^n \left(\frac{1}{m} \sum_{i=1}^m \frac{v_k^i}{\bar{V}} \right)^2 = \\ & \frac{1}{\bar{V}} \sum_{k=1}^n \bar{V}_k^2 = \bar{V} = \\ & \frac{1}{m^2 \bar{V}} \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^n \frac{v_k^i v_k^j}{\bar{V}} = \\ & \frac{1}{m^2 \bar{V}} \sum_{i=1}^m \sum_{j=1}^m (\bar{V}^i, \bar{V}^j) = \\ & \frac{1}{m^2 \bar{V}} \left(\sum_{i=1}^m (\bar{V}^i, \bar{V}^i) + 2 \sum_{1 \leq i < j \leq m} (\bar{V}^i, \bar{V}^j) \right) = \\ & \frac{1}{m^2 \bar{V}} (m+2C_m^2(\bar{V})) \end{aligned}$$

于是,易得本命题结论(2)。

(3)由相似度、中心点的定义及结论(2),得

$$\begin{aligned} (\bar{V}, V) &= \frac{1}{m} \sum_{i=1}^m (\bar{V}^i, V) = \frac{1}{m} \sum_{i=1}^m \sum_{k=1}^n \frac{v_k^i v_k}{\bar{V}} = \\ & \frac{1}{m} \sum_{k=1}^n \sum_{i=1}^m \frac{v_k^i v_k}{\bar{V}} = \sum_{k=1}^n \frac{v_k \bar{V}_k}{\bar{V}} = \\ & \bar{V} (\bar{V}, \bar{V}) = \sqrt{\frac{1+(m-1)(\bar{V})}{m}} (\bar{V}, \bar{V}) \end{aligned}$$

(4)添加偏好向量 \$v\$ 到向量集 \$V\$ 中,根据定义及结论(3)有

$$\begin{aligned} (\{v\}, V) &= \frac{1}{C_{m+1}^2} \left(\sum_{i=1}^m (\bar{V}^i, v) + \sum_{1 \leq i < j \leq m} (\bar{V}^i, \bar{V}^j) \right) = \\ & \frac{1}{C_{m+1}^2} (m(\bar{V}, v) + C_m^2(\bar{V})) = \\ & \frac{2}{m+1} \left(\sqrt{\frac{1+(m-1)(\bar{V})}{m}} (\bar{V}, \bar{V}) + \frac{m-1}{2} (\bar{V}) \right) \end{aligned}$$

当 \$m=1\$ 时,命题的结论(1) (2) (3) (4)显然成立,且由结论(1) (2) (3)和命题1的(3)可直接得如下推论1、推论2:

推论1 中心点 \$\bar{V}\$ 与 \$V\$ 的相似度满足: \$(\bar{V}, V) \geq \sqrt{(\bar{V})}\$

$$(\bar{V}) \geq (\bar{V}, V) \geq \bar{V}$$

推论2 对偏好向量 \$v\$, \$(v, V) \geq \gamma\$ 当且仅当 \$(v, \bar{V}) \geq \gamma \sqrt{\frac{m}{1+(m-1)(\bar{V})}}\$。

推论1指出,向量集或聚集的中心点与该集合的相似度 \$(\bar{V}, V)\$ 不低于向量集的一致性指标 \$(\bar{V})\$,也不低于向量集内的向量与该向量集的相似度的平均值 \$\bar{V}\$。因此,用该中心点作为向量集的代表(此时称之为向量集的偏好向量),比用集合内的某个成员更为合理。

定义5^[4] 聚集 \$C\$ 的偏好向量定义为该聚集的中心点经规范化后所得的向量,记为: \$\tilde{C} = \frac{\bar{C}}{\bar{C}}\$。群体的偏好向量定义为群体聚类后聚集的偏好向量加权平均并规范化后得到的向量。如果记 \$\tilde{\Omega} = \frac{1}{m} \sum_{i=1}^k \eta_i \tilde{C}_i\$, 则群体的偏好向量为 \$\tilde{\Omega} = \frac{\hat{\Omega}}{\hat{\Omega}}\$。

推论2表明,如果偏好向量 \$v\$ 与向量集 \$V\$ 的中心点 \$\bar{V}\$ 的相似度满足推论的条件(注意,该条件值不低于相聚阈值 \$\gamma\$),

则可保证 \$v\$ 与 \$V\$ 的相似度 \$(v, V) \geq \gamma\$。根据命题2, \$v\$ 添加到 \$V\$ 后,新的向量集仍然保持其一致性指标不小于阈值 \$\gamma\$。这样,命题2条件中的计算候选向量与向量集中的全部成员相似度的平均值,就转化为计算与中心点的相似度。

需要特别指出,如果群体成员向量全部规范化后再进行聚集计算,则本文的聚集中心点与文献[4]算法中聚集内各向量平均值所成的代表向量是相同的,但文献[4]仍然用给定的阈值作为相聚准则,从而导致启发式搜索算法所得的某些聚集的一致性指标较低,也导致群体或某些聚集的相对偏差指标较大,具体比较见本文的第4章。

3 群体聚类算法

根据前面的性质与命题,给出聚类算法如下:

步骤1 给出群体 \$V\$ 及判定阈值 \$\gamma\$;

步骤2 对 \$V\$ 中的偏好向量随机排序并规范化(2-范数为1),令聚集集合 \$C\$ 为空集;

步骤3 取 \$C\$ 中向量 \$v, -\{v\}\$。令临时集 \$T\$ 为空,聚集 \$C=\{v\}\$,聚集成员数 \$m=1\$,聚集一致性指标 \$\gamma=0\$,聚集中心点 \$c=v\$,向量与中心点 \$c\$ 的相聚阈值 \$\lambda=\gamma\$;

步骤4 如果 \$C\$ 为空,则转步骤9;

步骤5 取 \$C\$ 中向量 \$v, -\{v\}\$;

步骤6 如果 \$(v, c) < \lambda\$ (不满足相聚充要条件),则 \$T=\{v\}\$, \$T\$ 转步骤4;

步骤7 \$C+=\{v\}\$, \$C, m=m+1\$,迭代计算一致性指标、中心点 \$c\$ 和相聚阈值 \$\lambda\$ 如下:

$$\begin{aligned} & \gamma = \frac{2}{m} \left(\sqrt{\frac{1+(m-2)(\bar{V})}{m-1}} (\bar{V}, c) + \frac{m-2}{2} (\bar{V}) \right) \\ c &= \frac{(m-1)c+v}{m}, \lambda = \gamma \sqrt{\frac{m}{1+(m-1)(\bar{V})}} \end{aligned}$$

步骤8 转步骤4;

步骤9 \$C+=T\$;

步骤10 如果 \$T\$ 为空,停止计算并输出聚集集合 \$C\$;

步骤11 \$T=C\$ 转步骤3。

上述启发式算法需要说明几点:(1)命题1保证了步骤2中对群体 \$V\$ 的向量作规范化后不影响聚类结果。同时,计算向量相似度时均需要计算其范式值,故也不会增加本算法的计算工作量;(2)步骤7的迭代计算顺序中,必须先计算成员个数 \$m\$,再计算一致性指标 \$\gamma\$,最后计算聚集中心点 \$c\$ 和相聚阈值 \$\lambda\$;(3)聚集添加新的成员后,中心点、成员一致性指标、成员与中心点的相聚判断阈值均是迭代计算得到。其中,聚集中心点的迭代式 \$c = \frac{(m-1)c+v}{m}\$ 考虑了群体成员向量的规范性;(4)与一般的动态计算中心点的算法^[4]或比较聚集内各个点^[1-3]的算法不同,本算法的判定步骤中,只需要进行两个向量的算术运算,以及与成员个数无关的迭代计算工作。于是,在群体数较大时本算法将显著减少计算工作量。

4 测试结果与分析

本文主要以文献[4]的算法及实例作为比较,测试与分析聚类后的一致性指标、聚集个数及相对偏差指标。

4.1 一致性指标及聚集个数比较

取文献[4]中30个成员的群体实例。表1给出了当阈值 \$\gamma=\$

0.80 时两种算法的聚类结果。

表 1 阈值 $\gamma=0.80$ 时的聚类结果

本文算法			文献[4]的算法		
聚集序号	成员个数	一致性指标	聚集序号	成员个数	一致性指标
1	15	0.86	1	9	0.71
2	3	0.86	2	4	0.61
3	5	0.86	3	5	0.91
4	5	0.90	4	4	0.90
5	1	0.00	5	2	0.87
6	1	0.00	6	3	0.89
			7	1	0.00
			8	2	0.97
群体	30	0.81	群体	30	0.78

可以看出,本文算法所得各个聚集的一致性指标均高于 0.80 (一个成员的聚集除外),群体的一致性指标也较好。而文献[4]算法的聚类结果中,一致性指标差异较大:有低于阈值的(如 0.61),也有高于阈值的(如 0.97),群体的一致性指标也较低。

表 2 是用随机方法产生的 5 个属性且属性值在 0-1 之间的成员矢量,当个数从 100 到 1 000 且阈值固定为 0.80 时,两种聚类算法所得的聚集个数比较。可以看出,随着成员个数的增加,本文算法所得的聚集数缓慢增加,显示出较好的算法可伸缩性能。

表 2 阈值 $\gamma=0.80$ 时的聚类结果

成员个数	聚集个数	
	本文算法	文献[4]的算法
100	9	13
200	12	12
300	12	14
400	13	16
500	13	15
600	11	19
700	12	20
800	14	20
900	12	22
1 000	14	22

4.2 相对偏差指标比较

仍用前面 30 个成员的群体实例作为聚类对象。当阈值从 0 到 1.0、步长为 0.05 时,两种算法聚类结果中的相对偏差指标见图 1。图 2 则是固定阈值为 0.80 时,随机产生的成员个数从 100 变化到 1000 时两种算法的群体相对偏差指标。

图中显示,两种情况下本文的算法均有较低的群体相对偏差指标。从图 1 中还可看出,当阈值从 0.55 趋近 1 时,群体的相对偏差指标也同步降低,符合阈值趋 1 迫使相似度也趋 1 从而相对偏差趋 0 的实际。图 2 则显示了本文算法的群体相对偏差指标只有文献[4]的 50%左右,表明本文算法得出的聚集中

各成员的“比较相似”的聚类特征。

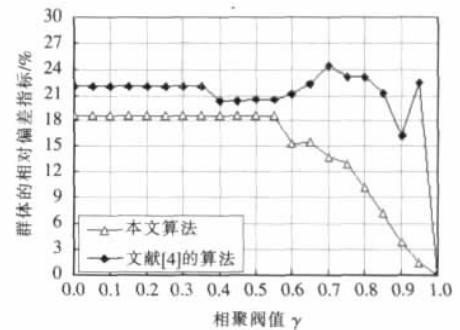


图 1 阈值变化时的群体相对偏差指标

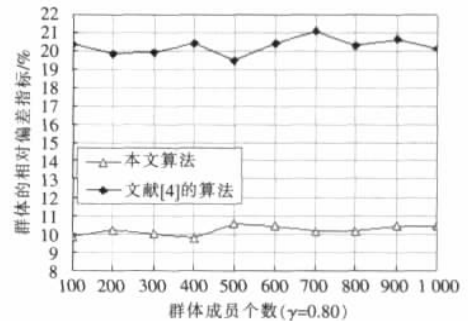


图 2 成员数变化时的群体相对偏差指标

5 结语

本文的主要工作总结为:(1)提出了偏好矢量与矢量集的相似度、聚集的中心点、聚集及群体的偏差指标、聚集及群体的相对偏差指标等概念;(2)证明了几个关于相似度、一致性指标等有用的性质、命题和推论;(3)给出了保持聚集一致性的偏好矢量与聚集相聚的充分必要条件;(4)论述了聚集中心点作为聚集的偏好矢量的合理性;(5)应用偏差指标及相对偏差指标评估聚类结果;(6)给出了一个启发式迭代算法。最后,用测试实例说明,本文的算法有较好的群体一致性指标与算法可伸缩性、较低的群体相对偏差指标。(收稿日期:2007年1月)

参考文献:

- [1] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms [C]//Proceedings of the 10th International World Wide Web Conference 2001 285-295.
- [2] 邓爱林, 朱扬勇, 施伯乐. 基于项目评分预测的协同过滤推荐算法[J]. 软件学报, 2003, 14(9): 1621-1628.
- [3] 褚忠信, 赵骞, 黄海挺. 基于相似系数的 1 种算法在遥感图像分类中的应用[J]. 青岛海洋大学学报, 2003, 33(5): 791-794.
- [4] 徐选华, 陈晓红. 基于矢量空间的群体聚类方法研究[J]. 系统工程与电子技术, 2005, 27(6): 1034-1037.
- [5] 李洁明, 祁新娥. 统计学原理[M]. 2版. 上海: 复旦大学出版社, 1999.
- [6] 张贤达. 矩阵分析与应用[M]. 北京: 清华大学出版社, 2004.
- [7] Someya H, Yamamura M. Genetic algorithm with search area adaptation for the function optimization and its experimental analysis[C]//Proceedings of IEEE Seoul Neural Networks Council 2001 933-940.
- [8] Someya H, Yamamura M. A genetic algorithm for the floorplan design problem with search area adaptation along with searching stage[J]. T IEE Japan, 1999, 119-C(3): 393-403.
- [9] Michalewicz Z. Genetic algorithms+data structures evolution programs[M]. Berlin: Springer-Verlag, 1994.

(上接 43 页)

- [3] 王凌. 智能优化算法及其应用[M]. 北京: 清华大学出版社, 2001.
- [4] De Jong K A. An analysis of the behavior of a class of genetic adaptive systems[D]. University of Michigan, 1975.
- [5] 徐俊刚. 流程企业生产调度方法与应用研究[D]. 北京: 中国科学院软件研究所, 2003.
- [6] 杨晓梅, 曾建潮. 采用多个体交叉的遗传算法求解作业车间问题[J]. 计算机集成制造系统, 2004, 10(9): 1114-1119.