

GP、SAP 算法比较研究

吴晓军, 尹俊

(西北工业大学自动化学院, 陕西 西安 710072)

摘要:针对规划问题,引入了固定结构解的描述形式,指出其离散量与连续量混合的多峰值优化的特点。在此基础上提出了固定结构遗传规划算法(GP)、模拟退火规划算法(SAP),并进行了算法分析。最后通过实验对四个典型优化函数的优化进行了比较研究。研究与实验结果表明SAP算法综合考虑了结构优化与参数优化,具有收敛效率高、获得更优解概率大的特点;GP算法有利于结构优化,但不利于参数优化,具有收敛效率较低,获得更优解的概率较小的特点。

关键词:规划;遗传规划;模拟退火规划

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2011)02-0009-04

A Study of Comparisons of GP and SAP

WU Xiao-jun, YIN Jun

(School of Automation, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: Studies the description form of a fixed structure on programming problems characterized by multi-peak optimization in both combined discrete and continuous volume. Based on the idea, fixed structure genetic programming (GP) and simulated annealing programming (SAP) are proposed, analyzed and finally compared through experiments on optimization of four typical optimization functions. Research and experimental results show that the SAP algorithm which takes into account of both the structure optimization and parameter optimization has a convergence of high efficiency and a large probability of arriving at a better solution; The GP algorithm is conducive to structure optimization, but not to parameter optimization, and has a convergence of low efficiency and a small probability of obtaining a better solution.

Key words: programming; genetic programming; simulated annealing programming

0 引言

遗传规划(Genetic Programming, GP)^[1]是进化算法的一种。它把不同领域的问题归结为寻找满足给定约束的计算机程序发现问题(即Programming问题,以下简称规划问题),也就是在可能的程序空间中寻找最优或满意的计算机程序^[2]。目前,遗传规划已经成功地应用于预测分析^[3]、数据挖掘^[4]、系统建模^[5]等工程实践领域中。

GA-PSO混合规划算法将规划问题的层次型结构转换为具有固定长度的线性结构,最后通过遗传算法及PSO算法进行求解^[6]。这种固定结构的解的描述方法使得更多的优化算法,如模拟退火算法等应用于规划问题的求解成为可能。模拟退火算法(Simulated Annealing, SA)^[7]是一个以概率收敛于全局最优解的

优化算法,是近年来研究的热点,目前已有应用于规划问题求解的研究成果^[8,9]。

文中将规划问题描述为具有固定长度的线性结构,提出了固定结构GP算法、模拟退火规划算法等,并将两种算法分别针对四个不同特征的优化函数的优化进行了比较研究。

1 GP、SAP 算法

1.1 固定结构解的描述

在函数集 F 中分别加入取左、右子树的运算符LT、RT;终止符集 T 由常量与变量构成。所有规划问题的解均可以描述为指定深度的满二叉树,并通过遍历实现线性化^[6]。如图1,表达式树描述了公式: $y = a \ln(x_1) + b e^{(x_1 x_2)}$ 。

通过改造函数集 F ,可以将GP中所有层次型问题,通过一个完全二叉树^[10]来描述。设表达式树的深度为 d ,可以采用 $2^d - 1$ 个结点的静态表存储深度为 d 的二叉树,一般情况下通过前序遍历得到满二叉树的静态表。设 $D = 2^d - 1$,则可将解描述为一个 D 维向量,令 $d = 5$,则表达式可表示为表1的形式,文中将这

收稿日期:2010-07-26;修回日期:2010-10-26

基金项目:国家自然科学基金项目(60970054)

作者简介:吴晓军(1970-),男,副教授,研究方向为嵌入式系统、计算机网络与通信、模式识别、智能信息处理、工业过程控制以及复杂系统研究。

种固定大小的静态表结构简称为固定结构。

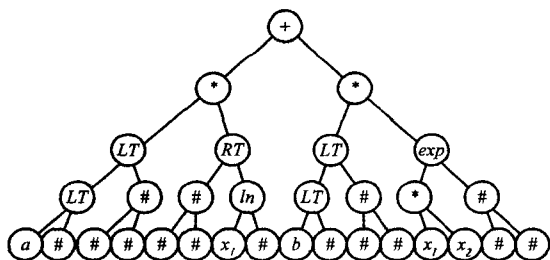


图 1 表达式的满二叉树表示

表 1 图 1 的固定长度静态表表示

位置	1	2	3	4	5	6	7	8
符号	+	*	LT	LT	a	#	#	#
位置	9	10	11	12	13	14	15	16
符号	#	RT	#	#	#	Ln	x ₁	#
位置	17	18	19	20	21	22	23	24
符号	*	LT	LT	B	#	#	#	#
位置	25	26	27	28	29	30	31	
符号	exp	*	x ₁	x ₂	#	#	#	

在静态表中,以函数的序号表示函数,以变量的序号表示变量,以浮点数表示常量。在生成解时,随机选取函数序号填入表达式串的函数位置,随机选取变量序号表示变量、随机产生浮点数表示常量,并填入表达式串的终止符位置,从而形成规划问题固定结构的解。

1.2 解的适应度

设对应描述解的字符串 S 的表达式为 f_s 。 x_i 为第 i 组样本, y_i 为训练样本中的观测值。则具有 n 组样本的串 S 的适应度表示为公式(1):

$$\text{fitness} = \sum_{i=1}^n [f_s(x_i) - y_i]^2 \quad (1)$$

1.3 固定结构 GP 算法

固定结构 GP 算法(文中仍简称 GP)利用遗传算法对固定结构的解进行优化,遗传操作与 GA-PSO 混合规划算法^[6]中描述的不同。

- (1)生成初始群体;
- (2)对每个解根据公式(1)计算解适应度;
- (3)遗传操作;
- (4)更新解的适应度;
- (5)反复执行(3)、(4),直到满足终止条件,输出最优解。

1.4 模拟退火规划算法 SAP

模拟退火算法的思想源于固体的退火过程:固体加热时内部粒子趋向于无序,内能增大,缓慢冷却时粒子又趋向有序,且在理想状态下每个温度都达到平衡,直到达内能最小的基态。根据 Metropolis 准则,粒子在温度 T 时趋于平衡的几率为 $p = e^{-\frac{\Delta E}{KT}}$,其中 ΔE 为内能改变量, T 为温度, K 为 Boltzmann 常数^[11]。利用模拟退火算法对规划问题求解称为模拟退火规划算法

(Simulated Annealing Programming, SAP),算法描述为:

- (1)初始化解,并设置温度 T_0 、迭代次数 L 、退火系数 α 和终止条件;
- (2)设解为 M_0 ,并根据公式(1)计算适应度 $f(M_0)$;
- (3)在 M_0 的邻域内产生新解 M ,并计算适应度改变量 $\Delta f = f(M) - f(M_0)$;
- (4)若 $\Delta f < 0$ 则接受 M ,否则按照 Metropolis 准则判断是否接受;
- (5)执行(3)~(4),并迭代 L 次;
- (6)降低温度,并重复(3)~(5),如果满足终止条件则结束程序,输出最优解。

1.5 算法分析

规划问题是在一个庞大的表达式空间上搜索与样本的最优匹配,它具有以下几个特点:

(1)规划问题的解中既存在描述结构的离散量,又存在描述参数的连续量。因此它是离散量与连续量的混合优化,是参数与结构相融合优化。

(2)搜索空间大。设函数集有 n_f 个函数,终止符集有 n_e 个变量,不考虑浮点型常量,则在深度为 d 的表达式中,其空间大小为:

$$S = 2^{(d-1)n_e} (2^{d-1} - 1)^{n_f} \quad (2)$$

如果考虑常量的优化,则搜索空间更大。

(3)多峰值优化。规划问题中相同的表达式能够描述为不同的形式,而每一种形式都可能形成一个峰值。如表达式中 $\cos(x)$ 、 $\sin(0.5\pi + x)$ 形式不同,但结果却相同,因此规划问题是一种多峰值优化。

GP 算法通过进化过程中遗传操作使得最优个体的基因片段转向更多的个体。结合规划问题,GP 算法在交叉操作的过程中,父代的结构片段向子代复制,从而形成新的描述结构,连续量不会发生变化。变异操作是对解的局部调整,函数及变量上变异对子代的结构产生影响,常数上的变异使得常数产生一种随机“跳变”,而常规 GP 算法并未针对这一特征进行改进。由此可见,GP 算法能够获得具有更好结构的新解,但对于连续的参数则随机搜索,缺乏局部优化的策略。

SAP 是一种全局优化算法,其运算可以分为降温与保温两个阶段。在降温阶段,算法以 Metropolis 准则接受新解,新解的接受意味着搜索的基点发生了变化,而该变化并非完全指向更优的方向,因此这种搜索策略有利于突破局部最优的限制,使其具有了全局优化的能力。对应于规划问题,能够实现解的结构与参数上的变换,并为保温阶段的局部搜索提供了基点。保温阶段则属于一种微调,是在搜索基点的领域内寻找可接受的解,类似于爬山算法。SAP 算法在解的结构调整的同时,兼顾了参数的优化,因此其搜索策略更为

全面。

由以上分析可知,GP 算法偏重于全局优化,但对于局部优化,尤其规划问题中参数优化不足;SAP 算法的两个阶段既考虑了全局优化,又兼顾了参数优化,较其它算法能够获得更优的解。

2 算法实验设计

文中由实验出发,选择 4 个典型的单峰函数和多峰函数^[12]作为优化函数,同时通过多次实验选择 GP、SAP 算法的最优参数。

2.1 优化函数

如下 4 个典型的函数,确定其定义域,并在 4 维空间上进行研究。

(1) Sphere 函数,单峰值函数。

$$f(x) = \sum_{i=1}^3 x_i^2, \quad -6 < x_i < 6$$

(2) Rosenbrock 函数,单峰值函数。

$$f(x) = \sum_{i=1}^3 (100(x_{i-1} - x_i^2)^2 + (x_i - 1)^2)$$

 $-2 < x_i < 2$

(3) Rastrigin 函数,多峰值函数。

$$f(x) = \sum_{i=1}^3 (x_i^2 - 10\cos(2\pi x_i) + 10)$$

 $-4 < x_i < 4$

(4) Griewank 函数,多峰值函数。

$$f(x) = \frac{1}{4000} \sum_{i=1}^3 x_i^2 - \prod_{i=1}^3 \cos\left[\frac{x_i}{\sqrt{i}}\right] + 1$$

 $-60 < x_i < 60$

2.2 样本数据生成

对以上 4 个典型的优化函数,在设定的定义域内随机均匀选取 200 组样本数据。通过预处理删除相同或接近的样本,最后保留 100 组,并计算函数值,得到样本集合。

$$\{(x_{1i}, x_{2i}, x_{3i}, y_i) \mid i = 1, 2, \dots, 100\} \tag{3}$$

2.3 结果分析

多次重复实验所获得的解的离散性表示在多次试验中解偏离平均值的程度,在算法收敛效率接近的状态下能够反映算法的稳定性。这里采用标准差度量解的离散性,标准差越小表示算法的稳定性越高。

误差与解的分布性是指在本试验中,指定误差控制范围内多次重复实验得到的解的次数,该指标反映了算法获得更优解的概率,文中重点对比了两种算法在 20 次重复试验中得到 20% 误差线以内的解出现的次数。

2.4 仿真实验

针对不同的优化函数,通过多次实验确定 GP 和

SAP 两种算法的参数,并对 4 组样本数据进行函数建模(见表 2)。每种算法重复实验 20 次,每隔 5s 输出一次最优适应度。

表 2 为实验选择的参数

函数名称	GP		SAP		
	规模	交叉/变异	T ₀	a	L
Sphere	100	10/20	200	0.95	100n
Rosenbrock	100	30/20	150	0.95	100n
Rastrigin	100	10/40	210	0.95	100n
Griewank	100	60/20	240	0.95	100n

3 实验结果及分析

3.1 平均适应度

图 2 是 4 个优化函数对应两种优化算法的平均适应度曲线,其中横坐标为迭代的时间点,为了便于观察,时间起点一般由 10 时刻开始,纵坐标为误差,这两者均为相对的概念。在四个函数的优化中,SAP 算法在优化初期及后期,尤其后期表现出了良好收敛特性,且收敛趋势明显,从而验证了 SAP 算法全局与局部兼顾的优化思想。

3.2 解的分布

图 3 是四个典型函数对应两种优化算法的解的分布曲线。其中横轴为解的误差的相对度量,纵轴为该误差内解出现的次数,A 线为 20% 误差线。对 Sphere,误差控制在 A 线以内的实验,SAP 为 17,GP 为 15;对 Rosenbrock,误差控制在 A 线以内的实验,SAP 为 12,GP 为 5;对 Rastrigin,误差控制在 A 线以内的实验,SAP 为 8 次,GP 为 1 次;对 Griewank,误差控制在 A 线以内的实验,SAP 为 20 次,GP 为 4 次。由以上数据以及图 3,SAP 算法具有更好的获得最优解的能力。

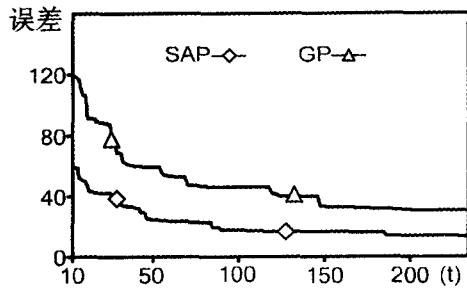
3.3 算法比较

通过算法分析与实验结果显示:①SAP 算法在规划问题的求解中收敛性优势明显,四种优化函数对于算法的收敛效率影响不大;②通过误差与解的分布对比实验显示,SAP 算法具有较好的获得更优解的概率,GP 算法次之。

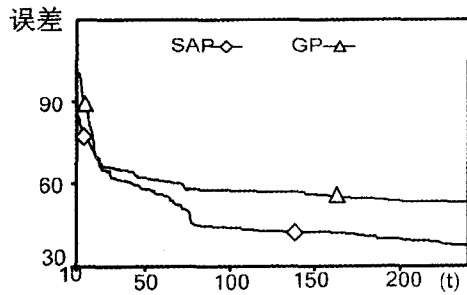
4 结束语

通过规划问题固定结构解的描述以及适应度的计算,总结了固定结构解的规划问题的特点,并结合遗传算法、模拟退火算法,提出了固定结构 GP、SAP 算法,在此基础上对两种算法进行了分析,得出:①模拟退火规划算法在全局优化的基础上兼顾局部优化,对规划问题的求解具有较强的针对性;②固定结构遗传规划算法有利于全局搜索,但对参数的优化能力不足。

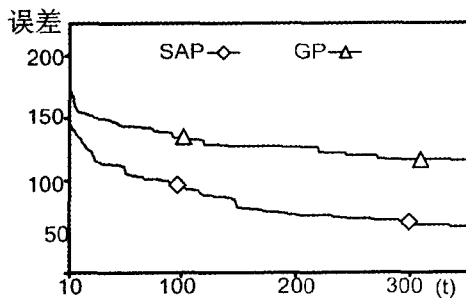
在对比实验中,文中从最优解收敛效率、解的分布



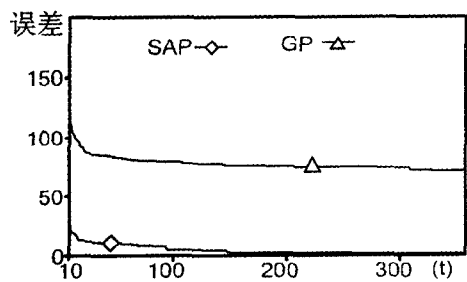
a. Sphere 函数



b. Rosenbrock 函数



c. Rastrigin 函数

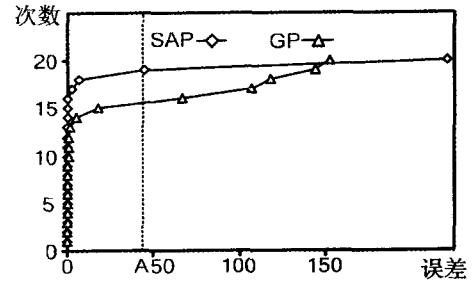


d. Griewank 函数

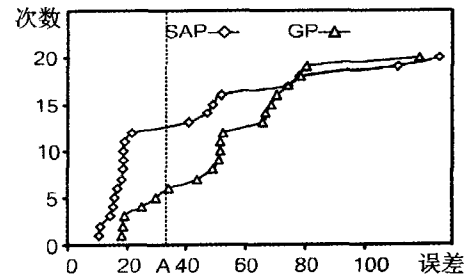
图 2 优化函数的平均适应度收敛曲线

两个方面比较了两种算法对四组典型函数的优化。结果表明模拟退火算法在解的收敛效率方面明显优于 GP 算法;在误差与解的分布试验中, SAP 算法获得最优解的可能更大,并且优势明显。

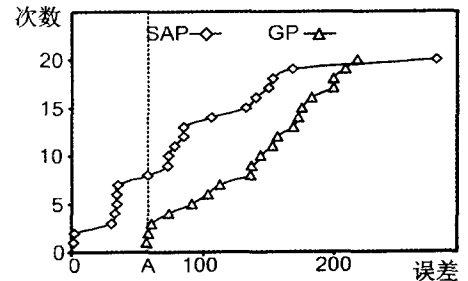
文中实验结果与算法分析结论一致。同时,在研究中发现 SAP 算法在规划问题的求解中具有明显的优势,且多次实验获得更优解的概率较高,由此,进一步设想,采用多点并行模拟退火算法将极大地改善规划算法的性能。



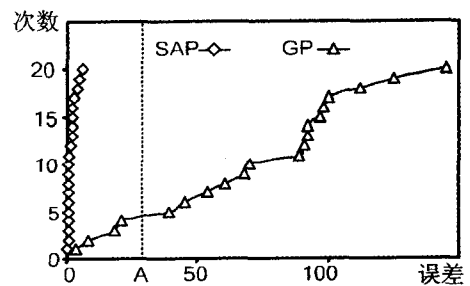
a. Sphere 函数



b. Rosenbrock 函数



c. Rastrigin 函数



d. Griewank 函数

图 3 误差与解的分布曲线

参考文献:

- [1] Koza J R. Genetic programming: on the programming of computers by means of natural selection [M]. Cambridge, MA: MIT Press, 1992.
- [2] Koza J R. Automatic Creation of Human - Competitive Programs and Controllers by Means of Genetic programming [J]. Genetic Programming and Evolvable Machines, 2000, 1:124-164.
- [3] Etemadi H, Rostamy A A A, Dehkordi H F. A Genetic Pro-

(下转第 17 页)

应商进行互连时产生的由两个虚拟活动组成的视图; 供应商仅与生产商进行互连, 因此创建一个视图。表 2 是合作伙伴的公开视图工作流。

表 2 伙伴工作流的公开视图

客户	生产商		供应商
对生产商的视图	对客户的视图	对供应商的视图	对生产商的视图
发送订单 ↓ 收到通知 ↓ 收到商品 ↓ 收到发票 ↓ 发送货款	收到订单 ↓ 发送通知 ↓ 发送商品 ↓ 发送发票 ↓ 收到货款	发送订单 ↓ 收到商品	收到订单 ↓ 发送商品

3 结束语

文章提出了一种基于视图的方法, 以此为基础来研究跨企业的动态协同中出现的问题。跨企业动态协同过程中必须考虑的问题包括对工作流的柔性支持、隐私保护原则及已建立的工作流的保护, 为满足上面的要求, 文中提出的基于视图的方法包括三个步骤: 工作流抽象与广播; 工作流匹配与互连; 工作流合作与监控。提出的方法考虑了工作流和资源的部分可见性, 使得企业能够保持其内部工作流隐私和安全性所需的水平, 从而把工作流的可见性降低到合作所需的最低水平。另外, 由于跨企业工作流可以直接采用参与协同的工作流而不必改变其内部结构, 从而为参与企业提供最大程度的灵活性。

参考文献:

[1] 范玉顺. 工作流管理技术基础——实现企业业务过程重组、过程管理和业务过程自动化的核心技术[M]. 北京: 清华大学出版社; 柏林: 施普林格出版社, 2001.

(上接第 12 页)

gramming Model for Bankruptcy Prediction: Empirical Evidence from Iran [M]. Expert Systems with Application (36), 2009, 2, 3199–3207.

[4] Guo Hong, Zhang Qing, Nandi, Asoke K. Feature extraction and dimensionality reduction by genetic programming based on the Fisher criterion [M]. Expert Systems (5), 2008, 11 (5): 444–459.

[5] 曹宏庆, 康立山, 陈毓屏, 等. 常微分方程组并行演化建模的实验研究[J]. 软件学报, 2003, 14 (3): 443–450.

[6] 吴晓军, 薛惠锋, 李 慇, 等. GA-PSO 混合规划算法[J]. 西北大学学报, 2005, 35(1): 39–43.

[7] Kirkpatrick S, Gelatt C D, Vecchi M P. Optimization by simulated annealing [J]. Science, 1983, 220(5): 671–680.

[2] 曹健辉, 李 峰. 柔性工作流动态建模机制的研究与设计[J]. 计算机技术与发展, 2009, 19(10): 47–50.

[3] 王 浩, 武 凌, 张 海, 等. 基于 Ontology 的工作流知识管理系统的研究[J]. 计算机技术与发展, 2008, 18 (6): 23–27.

[4] vander Aalst W M P. Process-oriented architectures for electronic commerce and inter-organizational workflow [J]. Information Systems, 2000, 25 (1): 43–69.

[5] Schmidt R. Web services based architectures to support dynamic inter-organizational business processes [C]//ICWS- Europe. [s. l.]: [s. n.], 2003: 123–136.

[6] 张月菊, 王 涛, 林 拉. 跨组织工作流集成中间件语义转换部件研究[J]. 计算机技术与发展, 2010, 20(3): 92–95.

[7] 唐 达, 钱会敏, 吴占洲. 基于角色的动态跨组织商业过程的工作流研究[J]. 小型微型计算机系统, 2006, 27 (5): 945–949.

[8] Grefen P, Aberer K, Hoffner Y, et al. CrossFlow: cross-organizational workflow management in dynamic virtual enterprises [J]. Int. Journal of Computer Systems Science & Engineering, 2000, 15(5): 277–290.

[9] 刘慧敏, 王刊良. 组织间协同工作流的交互模式研究[J]. 现代管理科学, 2009(10): 53–54.

[10] Grefen P, Aberer K, Hoffer Y, et al. Crossflow: Cross-Organizational Workflow Management for Service Outsourcing in Dynamic Virtual Enterprises[J]. IEEE Data Eng. Bull., 2001, 24(1): 52–57.

[11] Chebbi I, Dustdar S, Tata S. The view-based approach to dynamic inter-organizational workflow cooperation[J]. Data & Knowledge Engineering, 2006, 56: 139–173.

[12] Tata S, Klai K, M'bareck M O A. CoopFlow- A bottom-up approach to workflow cooperation for short-term virtual enterprises[J]. IEEE Transactions on Services Computing, 2008, 1(4): 214–228.

[8] 李少波, 徐立章, 胡建军. 基于遗传编程的可持续模拟退火算法及应用[J]. 系统仿真学报, 2009, 21 (7): 1908–1911.

[9] 胡为成, 王本年, 程转流. 基于模拟退火算法的遗传程序设计方法[J]. 计算机工程与设计, 2006, 27 (23): 4437–4439.

[10] 云庆夏. 进化算法[M]. 北京: 冶金工业出版社, 2000.

[11] 康立山. 非数值并行算法——模拟退火算法[M]. 北京: 科学出版社, 1994.

[12] Angeline P J. Evolutionary optimization versus particle swarm optimization: philosophy and performance differences[J]. In: Proceedings of 7th Annual Conference on Evolutionary Programming, Germany, 1998: 601–610.