

遗传算法研究综述*

葛继科¹, 邱玉辉¹, 吴春明¹, 蒲国林²

(1. 西南大学 计算机与信息科学学院, 重庆 400715; 2. 四川文理学院 计算机科学系, 四川 达州 635000)

摘要: 介绍了遗传算法的基本工作原理和主要特点, 讨论了遗传算法的理论、技术、存在问题及改进方法, 概述了遗传算法的常见应用领域, 分析了近五年国内对遗传算法的研究现状。最后, 进一步探讨了遗传算法的未来研究方向。

关键词: 遗传算法; 算子; 优化; 收敛性

中图分类号: TP301.6 文献标志码: A 文章编号: 1001-3695(2008)10-2911-06

Summary of genetic algorithms research

GE Ji-ke¹, QIU Yu-hui¹, WU Chun-ming¹, PU Guo-lin²

(1. School of Computer & Information Science, Southwest University, Chongqing 400715, China; 2. Dept. of Computer Science, Sichuan University of Arts & Science, Dazhou Sichuan 635000, China)

Abstract: This paper introduced the history, basic principle and main characters of genetic algorithms, discussed the theory, technology, limitation and improving measures about genetic algorithm. Then summarized the implementation techniques and applications of genetic algorithms, analyzed the research state of genetic algorithms in China during the past five years, and pointed out the genetic algorithms' research directions in the future.

Key words: genetic algorithm(GA); operator; optimization; convergence

遗传算法是由美国 Michigan 大学的 Holland 教授于 1969 年提出, 后经 DeJong、Goldberg 等人归纳总结所形成的一类模拟进化算法^[1-3]。它来源于达尔文的进化论、魏茨曼的物种选择学说和孟德尔的群体遗传学说。遗传算法是模拟自然界生物进化过程与机制求解极值问题的一类自组织、自适应人工智能技术^[4], 其基本思想是模拟自然界遗传机制和生物进化论而形成的一种过程搜索最优解的算法, 具有坚实的生物学基础; 它提供从智能生成过程观点对生物智能的模拟, 具有鲜明的认知学意义; 它适合于无表达或有表达的任何类函数, 具有可实现的并行计算行为; 它能解决任何种类实际问题, 具有广泛的应用价值。因此, 遗传算法广泛应用于自动控制、计算科学、模式识别、工程设计、智能故障诊断、管理科学和社会科学等领域, 适用于解决复杂的非线性和多维空间寻优问题。

虽然遗传算法在许多领域中都有成功的应用, 但其自身也存在不足, 如局部搜索能力差、存在未成熟收敛和随机游走等现象, 导致算法的收敛性能差, 需要很长时间才能找到最优解等问题。这些不足阻碍了遗传算法的推广应用。如何改善遗传算法的搜索能力和提高算法的收敛速度, 使其更好地应用于实际问题的解决中, 是各国研究者一直探索的主要课题。

1 遗传算法的执行过程及特点

1.1 遗传算法的执行过程

遗传算法作为一种自适应全局优化搜索算法, 使用二进制遗传编码, 即等位基因 $\chi = \{0, 1\}$, 个体空间 $H_L = \{0, 1\}^L$, 且繁殖分为交叉与变异两个独立的步骤进行。其基本执行过程如

下^[5]:

a) 初始化。确定种群规模 N 、交叉概率 P_c 、变异概率 P_m 和置终止进化准则; 随机生成 N 个个体作为初始种群 $X(0)$; 置进化代数计数器 $t = 0$ 。

b) 个体评价。计算或估价 $X(t)$ 中各个体的适应度。

c) 种群进化。

(a) 选择(母体)。从 $X(t)$ 中运用选择算子选择出 $M/2$ 对母体 ($M = N$)。

(b) 交叉。对所选择的 $M/2$ 对母体, 依概率 P_c 执行交叉形成 M 个中间个体。

(c) 变异。对 M 个中间个体分别独立依概率 P_m 执行变异, 形成 M 个候选个体。

(d) 选择(子代)。从上述所形成的 M 个候选个体中依适应度选择出 N 个个体组成新一代种群 $X(t+1)$ 。

d) 终止检验。如已满足终止准则, 则输出 $X(t+1)$ 中具有最大适应度的个体作为最优解, 终止计算; 否则置 $t = t+1$ 并转 c)。

1.2 遗传算法的特点

遗传算法利用了生物进化和遗传的思想。它不同于枚举法、启发式算法、搜索算法等传统的优化方法, 具有如下特点:

a) 自组织、自适应和智能性。遗传算法消除了算法设计中的一个最大障碍, 即需要事先描述问题的全部特点, 并说明针对问题的不同特点算法应采取的措施。因此, 它可用来解决复杂的非结构化问题, 具有很强的鲁棒性。

b) 直接处理的对象是参数编码集, 而不是问题参数本身。

收稿日期: 2008-01-12; 修回日期: 2008-03-26 基金项目: 西南大学研究生科技创新基金资助项目(2006011)

作者简介: 葛继科(1977-), 男, 河南濮阳人, 博士研究生, 主要研究方向为人工智能、语义网络(gjki@swu.edu.cn); 邱玉辉(1938-), 男, 教授, 博导, 主要研究方向为人工智能、语义网络; 吴春明(1972-), 男, 博士研究生, 主要研究方向为网络技术; 蒲国林(1971-), 男, 博士研究生, 主要研究方向为人工智能。

c) 搜索过程中使用的是基于目标函数值的评价信息, 搜索过程既不受优化函数连续性的约束, 也没有优化函数必须可导的要求。

d) 易于并行化, 可降低由于使用超强计算机硬件所带来的昂贵费用。

e) 基本思想简单, 运行方式和实现步骤规范, 便于具体使用。

2 遗传算法的理论研究

2.1 编码表示

在许多问题求解中, 编码是遗传算法中首要解决的问题, 对算法的性能有很重要的影响。Holland 提出的二进制编码是遗传算法中最常用的一种编码方法, 它采用最小字符编码原则, 编/解码操作简单易行, 利于交叉、变异操作的实现, 也可以采用模式定理对算法进行理论分析^[11]。但二进制编码用于多维、高精度数值问题优化时, 不能很好地克服连续函数离散化时的映射误差; 不能直接反映问题的固有结构, 精度不高, 并且个体长度大、占用内存多。针对二进制编码存在的不足, 人们提出了多种改进方法, 比较典型的有以下几种:

a) 格雷码编码。为了克服二进制编码在连续函数离散化时存在的不足, 人们提出了用格雷码进行编码的方法, 它是二进制编码的变形^[6]。假设有一个二进制编码为 $X = x_m x_{m-1} \dots x_2 x_1$, 其对应的格雷码为 $Y = y_m y_{m-1} \dots y_2 y_1$, 则

$$\begin{cases} y_m = x_m & i = m-1, m-2, \dots, 1 \\ y_i = x_{i+1} \oplus x_i \end{cases}$$

格雷码不仅具有二进制编码的一些优点, 而且能够提高遗传算法的局部搜索能力。

b) 实数编码。该方法适合于遗传算法中表示范围较大的数, 使遗传算法更接近问题空间, 避免了编码和解码的过程。它便于较大空间的遗传搜索, 提高了遗传算法的精度要求^[6]; 便于设计专门问题的遗传算子; 便于算法与经典优化方法的混合作用, 改善了遗传算法的计算复杂性, 提高了运算效率。

c) 十进制编码。该方法利用十进制编码控制参数, 缓解了“组合爆炸”和遗传算法的早熟收敛问题^[7]。

d) 非数值编码。染色体编码串中的基因值取一个仅有代码含义而无数值含义的符号集, 这些符号可以是数字也可以是字符^[8]。非数值编码的优点是在遗传算法中可以利用所求问题的专门知识及相关算法。对于非数值编码, 问题的解和染色体的编码要注意染色体的可行性、染色体的合法性和映射的惟一性。

2.2 适应度函数

在遗传算法中, 适应度是描述个体性能的主要指标, 根据适应度的大小对个体进行优胜劣汰。对于求解有约束的优化问题时, 一般采用罚函数方法将目标函数和约束条件建立一个无约束的优化目标函数; 然后再将目标函数作适当处理, 建立适合遗传算法的适应度函数。将目标函数转换成适应度函数一般应遵循两个原则: 适应度必须非负; 优化过程中目标函数的变化方向应与群体进化过程中适应度函数变化方向一致^[9]。在使用遗传算法求解具体问题时, 适应度函数的选择对算法的收敛性以及收敛速度的影响较大, 针对不同的问题需根据经验或算法来确定相应的参数。

何新贵等人在对遗传算法的研究过程中, 考虑函数在搜索点的函数值及其变化率, 并将该信息加入适应度函数, 使得按概率选择的染色体不但具有较小的函数值, 而且具有较大的函数变化率值^[10]。实验结果表明, 这类方法的收敛速度明显高于标准的遗传算法。陶卿等人提出基于约束区域神经网络的动态遗传算法^[11], 将遗传算法的全局搜索和约束区域神经网络模型的局部搜索结合起来; 利用动态遗传算法确定神经网络模型的初始点, 同时使用神经网络确定动态遗传算法的适应度函数。

2.3 遗传算子

在遗传算法中通过一系列算子来决定后代, 算子对当前群体中选定的成员进行重组和变异。

1) 选择算子 选择操作通过适应度选择优质个体而抛弃劣质个体, 体现了“适者生存”的原理。Potts 等人概括了 23 种选择方法^[12]。常见的选择操作主要有以下几种:

a) 轮盘赌选择。选择某假设的概率是通过这个假设的适应度与当前群体中其他成员的适应度的比值而得到。此方法是基于概率选择的, 存在统计误差, 因此可以结合最优保存策略以保证当前适应度最优的个体能够进化到下一代而不被遗传操作的随机性破坏, 保证算法的收敛性。

b) 排序选择。对个体适应值取正值或负值以及个体适应度之间的数值差异程度无特殊要求, 对群体中的所有个体按其适应度大小进行排序, 根据排序来分配各个体被选中的概率。

c) 最优个体保存。父代群体中的最优个体直接进入子代群体中。该方法可保证在遗传过程中所得到的个体不会被交叉和变异操作所破坏, 它是遗传算法收敛性的一个重要保证条件; 它也容易使得局部最优个体不易被淘汰, 从而使算法的全局搜索能力变强。

d) 随机联赛选择。每次选取 N 个个体中适应度最高的个体遗传到下一代群体中。具体操作如下: 从群体中随机选取 N 个个体进行适应度大小比较, 将其中适应度最高的个体遗传到下一代群体中; 将上述过程重复执行 M (为群体大小) 次, 则可得下一代群体。

2) 交叉算子 交叉是指对两个相互交叉的染色体按某种方式相互交换其部分基因, 从而形成两个新的个体。它是产生新个体的主要方法, 决定了遗传算法的全局搜索能力, 在遗传算法中起关键作用。Potts 等人概括了 17 种交叉方法^[12]。几种常用的适用于二进制编码或实数编码方式的交叉算子如下:

a) 单点交叉。在个体编码串中随机设置一个交叉点后在该点相互交换两个配对个体的部分基因。

b) 两点交叉。在相互配对的两个个体编码串中随机设置两个交叉点, 并交换两个交叉点之间的部分基因。

c) 均匀交叉。两个相互配对个体的每一位基因都以相同的概率进行交换, 从而形成两个新个体。

d) 算术交叉。由两个个体的线性组合而产生出新的个体。

3) 变异算子 变异是指将个体染色体编码串中的某些基因座上的基因值用该基因座的其他等位基因来替换, 从而形成一个新的个体。它是产生新个体的辅助方法, 决定了遗传算法的局部搜索能力^[12]。变异算子与交叉算子相互配合, 可以共同完成对搜索空间的全局搜索和局部搜索, 从而使得遗传算法以良好的搜索性能完成最优化问题的寻优过程。在遗传算法

中使用变异算子主要有以下两个目的: 改善遗传算法的局部搜索能力; 维持群体的多样性, 防止出现早熟现象。下面是几种常用的变异操作:

a) 基本位变异。对个体编码串以变异概率 P 随机指定某一位或某几位基因进行变异操作。

b) 均匀变异(一致变异)。分别用符合某一范围内均匀分布的随机数, 以某一较小的概率来替换个体编码串中各个基因座上的原有基因值。均匀变异操作特别适合应用于遗传算法的初期运行阶段, 它使得搜索点可以在整个搜索空间内自由地移动, 从而可以增加群体的多样性, 使算法能够处理更多的模式。

c) 二元变异。需要两条染色体参与, 通过二元变异操作后生成两条新个体中的各个基因分别取原染色体对应基因值的同或/异或。它改变了传统的变异方式, 有效地克服了早熟收敛, 提高了遗传算法的优化速度^[13]。

d) 高斯变异。在进行变异时用一个均值为 μ 方差为 σ^2 的正态分布的一个随机数来替换原有基因值。其操作过程与均匀变异类似。

2.4 参数选择

遗传算法的参数选择一般包括群体规模、收敛判据、杂交概率和变异概率等。参数选择关系到遗传算法的精度、可靠性和计算时间等诸多因素, 并且影响到结果的质量和系统性能。因此, 在遗传算法中参数选择的研究非常重要。

在标准的遗传算法中经常采用经验对参数进行估计, 这将带来很大的盲目性从而影响算法的全局最优性和收敛性, 人们意识到这些参数应该随着遗传进化而自适应变化。基于这一观点, Davis 提出了自适应算子概率方法^[14], 即用自适应机制把算子概率与算子产生的个体表示适应性相结合, 高适应性值被分配高算子概率。Whitley 等人提出了自适应突变策略与一对父串间的 Hamming 距离成反比的观点^[15], 结果显示能有效地保持基因的多样性。张良杰等人通过引入 i 位改进子空间概念, 采用模糊推理技术确定选取突变概率的一般性原则^[16]。文献[17]中用模糊规则对选择概率和变异概率进行控制, 在线改变其值。相应的算例表明, 这种方法有较好的性能。

2.5 收敛性分析

遗传算法的收敛性通常是指遗传算法所生成的迭代种群(或其分布)收敛到某一稳定状态(或分布), 或其适应值函数的最大或平均值随迭代趋于优化问题的最优值。依不同的研究方法及其所应用的数学工具, 收敛性分析可分为 Vose-Liepins 模型、Markov 链模型和公理化模型等。

1) Vose-Liepins 模型

该模型是 Vose 和 Liepins 提出来的, 它用两个矩阵算子分别刻画比例选择与组合算子(即杂交算子与变异算子的复合), 通过这两个算子不动点的存在性与稳定性来刻画遗传算法的渐近行为^[18]。

Vose-Liepins 模型只适用于简单遗传算法, 可应用于比例选择、单点杂交和单点变异等, 没有推广到更具实用性的其他遗传算法执行策略中, 如锦标赛选择、多点杂交等。另外, Vose-Liepins 模型不易处理变异、杂交概率随时间变化的情形, 其框架亦很难推广到描述一般非二进制或连续变量情形的遗传算法。

Vose-Liepins 模型虽然在种群规模无限的假设下可精确刻

画遗传算法, 但在有限规模情形下却只能描述遗传算法的平均形态。为了克服这个缺陷, Nix 和 Vose 结合 Vose-Liepins 模型与 Markov 链描述, 发展了遗传算法的一个精确 Markov 链模型^[19]。

2) Markov 链模型

遗传算法的马氏链模型主要有三种: 种群马氏模型、Vose 模型^[19]和 Cerf 扰动马氏链模型^[20]。

种群马氏链模型将遗传算法的种群迭代序列视为一个有限状态马氏链加以研究, 主要是运用种群马氏链转移概率矩阵的某些一般性质分析遗传算法的极限行为。

在 Vose 模型中, 种群的状态由一个概率向量表示, 概率向量的维数为所有可能个体的数目。当种群规模趋于无穷大时, 相对频率的极限就代表了每一个个体在种群中出现的概率。在无限种群规模假设下, 可以导出表示种群的概率向量的不动点及其稳定性, 从而导致对遗传算法极限行为的刻画。

R. Cerf 利用 Azencott 等人关于模拟退火的一系列工作, 将遗传算法看成一种特殊形式的广义模拟退火模型, 利用动力系统的随机扰动理论对遗传算法的极限行为及收敛速度进行了研究^[20]。尽管在 Cerf 模型中所研究的马氏链序列仍然是种群序列, 但研究方法与种群马氏链模型有差异, 所以称之为 Cerf 扰动马氏链模型。

用 Markov 链模型描述遗传算法虽然有直接、精确的优点, 但由于有限状态 Markov 链理论本身的限制, 与 Vose-Liepins 模型类似, 该模型只能用于描述通常的二进制或特殊的非二进制遗传算法。另外, 这类方法所得收敛性一般是指相应的 Markov 链趋于某一平稳分布。这与优化中通常所指的收敛性定义不同, 它并不保证遗传算法一定能收敛到问题的全局最优解。再者, 相应 Markov 链的状态数(转移矩阵的规模)通常很大, 这使得具体确定、分析转移矩阵的性态十分困难, 因而对于较大规模的遗传算法, Markov 链分析只能基于遍历性考察而得出相应遗传算法收敛性的某些“粗糙”结论。

3) 公理化模型

遗传算法的公理化方法基于对模拟进化计算操作的公理化描述, 应用各个相关操作的本质特征数来对算法的收敛性直接进行概率估计^[21]。这一方法不仅适用于包括非齐次、自适应等在内的各种遗传算法分析, 而且分析方法本身直观、清晰, 所导出的结论也对遗传算法的各种参数选择提供参考依据。这一模型具有重要的理论意义与实用价值。

文献[5]还通过详细估计常见选择算子与演化算子的选择压、选择强度、保持率、迁入率、迁出率等参数, 导出了一系列具有重要应用价值的遗传算法收敛性结果。公理化模型也可用于其他模拟演化算法的收敛性分析。

2.6 欺骗问题

欺骗问题是遗传算法研究中的一个热点。根据模式定理可知, 低阶、短距的优模式在遗传子代中将以指数级增长, 最终, 不同的最优模式相互组合, 从而得到最优解。但是, 对于欺骗问题, 复制算子受欺骗条件的“迷惑”, 使最优低阶模式在组合后形成非最优高阶模式, 从而使遗传算法偏离最优解。由于欺骗问题的存在, 许多问题被归结为遗传算法难题, 使遗传算法的应用受到一定的局限。目前遗传算法的欺骗问题研究主要集中在三个方面: a) 设计欺骗函数, 如 Goldberg 曾设计一个离散遗传算法的最小欺骗问题。b) 修改遗传算法以解决欺骗

函数的影响,如黄炎等人提出一种基于可调变异算子求解遗传算法中欺骗问题的方法,即在遗传搜索过程中通过改变变异算子的方向和概率求解遗传算法的欺骗问题^[22]。该方法能够在消除遗传算法中欺骗性条件的同时保持群体的多样性,使遗传算法可以顺利地收敛到全局最优解。c) 理解欺骗函数对遗传算法的影响,何军等人讨论了一类遗传算法求解完全欺骗性问题的平均计算时间^[23]。

2.7 并行遗传算法

并行算法的基本思想是将一个复杂的任务分解为多个较简单的子任务,然后将各子任务分别分配给多个处理器并行执行求解。例如, Zeigler 等人把高性能并行遗传算法应用到大型并行计算机^[24],这种分而治之的思想可以由不同的方法和途径实现,导致了各种不同类型的并行遗传算法。并行遗传算法可以分为四大类,即全局并行遗传算法、粗粒度并行遗传算法、细粒度并行遗传算法和混合并行遗传算法。

侯广坤等人讨论了并行遗传算法的迁移现象及群体规模估算模型,分析了迁移的过程,揭示了迁移的实质,并提出了在理想条件下的迁移计算模型^[25],基于迁移计算模型导出了粗粒度并行遗传算法进化质量估量模型。王洪燕等人对并行遗传算法的研究现状进行了详细的介绍^[26]。

3 遗传算法的应用

遗传算法提供了一种求解复杂系统优化问题的通用框架,它不依赖于问题的具体领域,广泛应用于多种学科领域。

3.1 函数优化

函数优化是遗传算法的经典应用领域,也是对遗传算法进行性能评价的常用算例。不同研究者构造出了各种各样的复杂形式的测试函数,有连续函数也有离散函数,有凸函数也有凹函数,有低维函数也有高维函数,有确定函数也有随机函数,有单峰值函数也有多峰值函数等。用这些几何特性各具特色的函数来评价遗传算法的性能,更能反映算法的本质效果。对于一些非线性、多模型、多目标的函数优化问题,用其他优化方法较难求解,遗传算法却可以方便地得到较好的结果^[27]。

3.2 组合优化

随着问题规模的增大,组合优化问题的搜索空间也急剧扩大,有时在目前的计算机上用枚举法很难或甚至不可能求出其精确的最优解。对这类复杂问题,应把主要精力放在寻求其满意解上,而遗传算法是寻求这种满意解的最佳工具之一。实践证明,遗传算法已经在求解旅行商问题、背包、装箱、布局优化、图形划分等各种具有 NP 难度的问题上得到成功的应用^[6,28~30]。

3.3 生产调度

在很多情况下,用常规方法建立的数学模型难以精确求解生产调度问题,即使经过一些简化之后可以进行求解,有时也会因简化太多而使得求解结果与实际目标相差甚远。一般情况下,在现实生产中主要是靠经验来进行调度。研究发现,遗传算法已成为解决复杂调度问题的有效工具,在单件生产车间调度、流水线生产车间调度、生产规划、任务分配等方面,遗传算法都得到了有效的应用^[8,31,32]。

3.4 自动控制

在自动控制领域中有很多与优化相关的问题需要求解,遗

传算法已在其中得到了初步的应用,并显示出良好的效果。例如用遗传算法进行航空控制系统的优化^[33]、使用遗传算法设计空间交汇控制器、基于遗传算法的模糊控制器的优化设计^[34]、基于遗传算法的参数辨识、基于遗传算法的模糊控制规则的学习,利用遗传算法进行人工神经网络的结构优化设计和权值学习等,都显示出了遗传算法在这些领域中应用的可能性。

3.5 机器人学

机器人是一类复杂的、难以精确建模的人工系统。由于遗传算法的起源来自于人工自适应系统的研究,机器人学理所当然地成为遗传算法的一个重要应用领域。遗传算法已经在移动机器人路径规划^[35]、关节机器人运动轨迹规划、机器人逆运动学求解、细胞机器人的结构优化和行为协调等方面进行了研究和应用^[36]。

3.6 图像处理

图像处理是计算机视觉中的一个重要研究领域。在图像处理过程中,如扫描、特征提取、图像分割等不可避免地会存在一些误差,从而影响图像的效果。如何使这些误差最小化是计算机视觉达到实用化的重要要求。遗传算法可用于图像处理中的优化计算,目前已在模式识别(包括汉字识别^[37])、图像恢复、图像边缘特征提取等方面得到了应用^[38,39]。

3.7 人工生命

人工生命是用计算机、机械等人工媒体模拟或构造出的具有自然生物系统特有行为的人造系统。自组织能力和自学习能力是人工生命的两大主要特征。人工生命与遗传算法有着密切的联系,基于遗传算法的进化模型是研究人工生命现象的重要基础理论。虽然人工生命的研究尚处于初期阶段,但遗传算法已在其进化模型、学习模型、行为模型、自组织模型等方面显示出了初步的应用能力,并且必将会得到更为深入的应用和发展。人工生命与遗传算法相辅相成,遗传算法为人工生命的研究提供一个有效的工具,人工生命的研究也必将促进遗传算法的进一步发展^[40]。

3.8 遗传编程

遗传编程(genetic programming, GP)采用树型结构表示计算机程序,运用遗传算法的思想,通过自动生成计算机程序来解决问题。虽然遗传编程的理论尚未成熟,应用也有一些限制,但它已成功地应用于人工智能、机器学习等领域^[41]。Koza 等人描述了基本的遗传编程方法并且给出了很多可以被 GP 成功学习的程序^[42~44]。

3.9 机器学习

学习能力是高级自适应系统所具备的能力之一,基于遗传算法的机器学习,特别是分类器系统,在很多领域中都得到了应用。遗传算法被用于学习模糊控制规则,可以更好地改进模糊系统的性能;基于遗传算法的机器学习不但可以用来调整人工神经网络的连接权,也可用于人工神经网络结构的优化设计^[45,46]。

3.10 数据挖掘

数据挖掘能够从大规模数据库中提取隐含的、未知的、有潜在应用价值的知识和规则。许多数据挖掘问题可以看做是搜索问题。其中,数据库可看做是搜索空间,挖掘算法可看做

是搜索策略。应用遗传算法在数据库中进行搜索,对随机产生的一组规则进行进化,直到数据库能被该组规则覆盖,从而挖掘出隐含在数据库中的规则^[47-49]。Sunil 已成功地开发了一个基于遗传算法的数据挖掘工具^[50],利用该工具对两个飞机失事的真实数据库进行了数据挖掘实验,结果表明遗传算法是进行数据挖掘的有效方法之一。

4 近五年国内遗传算法研究现状

在前人研究的基础上,本文以已发表论文为依据,重点考查了 2002—2006 年国内学者及工程技术人员在遗传算法方面的研究情况,如表 1 所示。该表数据的统计源是中国知网(CNKI)的“电子技术及信息科学类”中收录的中文核心期刊。

表 1 2002—2006 年国内遗传算法研究现状

年度	综述	函数优化	组合优化	生产调度	自动控制	机器人学	图像处理	人工生命	遗传编程	机器学习	数据挖掘
2002	6	99	104	23	19	9	18	3	3	16	6
2003	9	102	78	27	8	8	27	2	3	20	10
2004	4	127	93	39	21	8	28	1	0	17	15
2005	3	120	77	32	25	8	32	2	2	26	21
2006	1	94	57	28	12	7	24	2	1	16	12

表 1 分别从综述、函数优化、组合优化、生产调度、自动控制、机器人学、图像处理、人工生命、遗传编程、机器学习、数据挖掘 11 个方面对近五年国内发表在中文核心期刊上、与遗传算法有关的论文进行了分类。图 1 是根据表 1 生成的对比分析图。

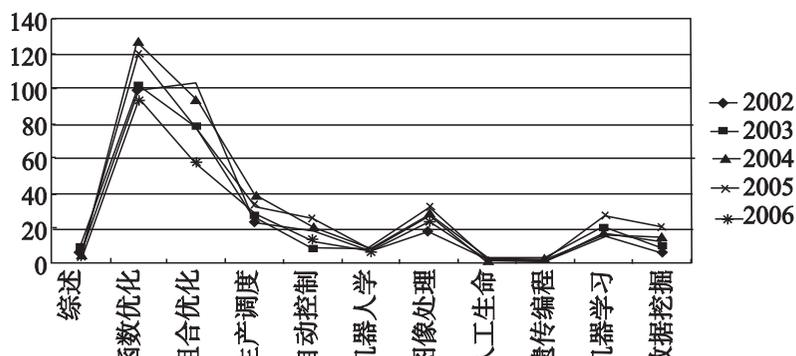


图 1 2002—2006 年国内遗传算法研究现状分析

通过对表 1 及图 1 的分析可以得出如下结论:

a) 针对遗传算法函数优化和组合优化方面进行研究的文章每年几乎都是最多的。这说明对遗传算法进行的研究更多的还是停留在理论层面,而生产调度及自动控制等实际应用领域的研究成果较少,有待进一步深入研究。

b) 总体来看,近五年国内对遗传算法的研究在逐年增长,但从 2005 年起有所回落,到 2006 年出现减少的现象。从这一现象来看,国内关于遗传算法的研究兴趣可能已达到饱和,理论研究已经比较成熟。另外,就目前研究现状来看,很难在原有的基础上得出更多新的成果。

c) 遗传算法在数据挖掘和机器学习领域进行的研究在研究成果中所占的比重逐年明显增长。这说明遗传算法在数据挖掘和机器学习领域的研究成为新的热点。虽然有效成果占总的研究成果的比例还很少,但这是一个可以进行深入研究的方向。

5 遗传算法的未来研究方向

遗传算法在各种问题的求解与应用中展现了其特点和魅力,同时也暴露出它在理论和应用上的诸多不足和缺陷。未来几年内遗传算法的研究重点可能集中在以下几方面:

a) 算法的数学基础,包括算法的收敛性、收敛速度估计、早熟机理的探索与预防、交叉算子的几何意义与统计解释、参数设置对算法的影响等方面。算法的收敛速度估计是当前特别值得研究和探讨的问题,它能从理论上对遗传算法的任何修正形式提供评判标准,从而指明改进算法性能的正确方向。

b) 遗传算法与优化技术的融合。对遗传算法的大范围群体搜索性能与快速收敛的局部优化方法进行混合,从而产生有效的全局优化方法。这种策略可从根本上提高遗传算法计算性能,对此可以进行大量的理论分析和实验。

c) 算法的改进与深化。根据具体应用领域对遗传算法进行改进与完善,仅泛泛地对一般问题进行研究是远远不够的。当前针对具体应用问题深化研究遗传算法是特别值得提倡的工作。

d) 算法的选择。基于实验研究的结论并不具有普遍意义上的指导作用,因此从数学的角度进行遗传算法的理论研究将对算法的合理选择提供理论指导。

e) 算法的并行化研究。遗传算法的群体适应度评价、随机搜索等特征使其具有明显的并行性。因此,设计各种并行执行策略、建立相应的并行化遗传算法的数学基础,是一项具有重要意义的工作。

6 结束语

遗传算法作为一种非确定性的拟自然算法,为复杂系统的优化提供了一种新的方法,实践证明其效果显著。尽管遗传算法在很多领域具有广泛的应用价值,但它仍存在一些问題,各国学者一直都在探索对遗传算法的改进及发展,以使遗传算法有更广泛的应用领域。

参考文献:

- [1] OLLAND J H. Adaptation in natural and artificial systems [M]. Ann Arbor: University of Michigan Press, 1975.
- [2] DeJONG K A. The analysis of the behavior of a class of genetic adaptive systems [D]. Ann Arbor: University of Michigan, 1975.
- [3] GOLDBERG D E. Genetic algorithms in search, optimization and machine learning [M]. Boston: Addison-Wesley Longman Press, 1989.
- [4] 张文修, 梁怡. 遗传算法的数学基础 [M]. 西安: 西安交通大学出版社, 2000.
- [5] 徐宗本, 张讲社, 郑亚林. 计算智能中的仿生学: 理论与算法 [M]. 北京: 科学出版社, 2003.
- [6] 周明, 孙树栋. 遗传算法原理及应用 [M]. 北京: 国防工业出版社, 1999.
- [7] 唐飞, 滕弘飞. 一种改进的遗传算法及其在布局优化中的应用 [J]. 软件学报, 1999, 10(10): 1096-1102.
- [8] 玄光男, 程润传. 遗传算法与工程设计 [M]. 北京: 科学出版社, 2000.
- [9] 邢文训, 谢金星. 现代优化计算方法 [M]. 北京: 清华大学出版社, 1999.
- [10] 何新贵, 梁久祯. 利用目标函数梯度的遗传算法 [J]. 软件学报, 2001, 12(7): 981-985.
- [11] 陶卿, 曹进德, 孙德敏, 等. 基于约束区域神经网络的动态遗传算法 [J]. 软件学报, 2001, 12(3): 462-467.
- [12] POTTS C J, TERRI D. The development and evaluation of an improved genetic algorithm based on migration and artificial selection [J]. IEEE Trans on Systems, Man, and Cybernetics, 1994, 24(1): 73-86.
- [13] 杨启文, 蒋静坪, 张国宏. 遗传算法优化速度的改进 [J]. 软件学报, 2001, 12(2): 270-275.

- [14] DAVIS L. Adaptive operator probability in genetic algorithms[C] // Proc of the 3rd International Conference on Genetic Algorithms. San Francisco: Morgan Kaufmann Publishers, 1989: 61-69.
- [15] WHITLEY D, STARKWEATHER D. Genetic II: a distributed genetic algorithms[J]. Journal of Experimental & Theoretical Artificial Intelligence, 1990, 7(7): 189-214.
- [16] 张良杰, 毛志宏, 李衍达. 遗传算法中突变算子的数学分析及改进策略[J]. 电子科学学刊, 1996, 18(6): 590-595.
- [17] LIU Li, CHEN Xue-yun. Reconfiguration of distribution networks based on fuzzy genetic algorithms[J]. Proceedings of the Chinese Society of Electrical Engineering, 2000, 20(2): 66-69.
- [18] VOSE M D, LIEPINS G E. Punctuated equilibria in genetic search [J]. Complex Systems, 1991, 5(1): 31-34.
- [19] NIX A E, VOSE M D. Modeling genetic algorithms with Markov chains[J]. Annals of Mathematics and Artificial Intelligence, 1992, 5(1): 79-88.
- [20] CERF R. Asymptotic convergence of a genetic algorithms [J]. Comptes Rendus de l'Académie des Sciences: Série 1, Mathématique, 1994, 319(33): 271-276.
- [21] GOLDBERG D E. Genetic and evolutionary algorithms come of age [J]. Communications of the ACM, 1994, 37(3): 113-119.
- [22] 黄炎, 蒋培, 王嘉松, 等. 基于可调变异算子求解遗传算法的欺骗问题[J]. 软件学报, 1999, 10(2): 216-219.
- [23] 何军, 黄厚宽, 康立山. 遗传算法求解完全欺骗问题的平均计算时间[J]. 计算机学报, 1999, 22(9): 999-1003.
- [24] ZEIGLER B P, KIM J. Asynchronous genetic algorithms on parallel computers[C] // Proc of the 5th International Conference on Genetic Algorithms. San Francisco: Morgan Kaufmann Publishers, 1991: 75-83.
- [25] 侯广坤, 骆江鹏. 一种理想并行遗传算法模型[J]. 软件学报, 1999, 10(5): 557-559.
- [26] 王洪燕, 杨敬安. 并行遗传算法研究进展[J]. 计算机科学, 1999, 26(6): 48-53.
- [27] 黄冀卓, 王湛, 马人乐. 一种新的求解约束多目标优化问题的遗传算法[J]. 计算机工程与应用, 2006, 42(23): 47-51.
- [28] 钱志勤, 滕弘飞, 孙治国. 人机交互的遗传算法及其在约束布局优化中的应用[J]. 计算机学报, 2001, 24(5): 553-558.
- [29] 王征应, 石冰心. 基于启发式遗传算法的 QoS 组播路由问题求解[J]. 计算机学报, 2001, 24(1): 55-61.
- [30] 赵振勇, 王力, 王保华, 等. 遗传算法改进策略的研究[J]. 计算机应用, 2006, 26(12): 189-191.
- [31] 李素粉, 朱云龙. 流水车间作业提前/拖期调度问题研究[J]. 计算机集成制造系统, 2006, 12(8): 1235-1240.
- [32] 饶运清, 严治雄, 张超勇, 等. 一种混合遗传算法在车间作业调度中的应用研究[J]. 机械科学与技术, 2006, 25(5): 584-587, 607.
- [33] 虞蕾, 赵红, 赵宗涛. 一种基于遗传算法的航迹优化方法[J]. 西北大学学报: 自然科学版, 2006, 36(2): 205-208, 213.
- [34] 李辉, 韩红, 韩崇昭, 等. 基于遗传算法的模糊逻辑控制器优化设计[J]. 西安交通大学学报, 2002, 36(4): 385-389.
- [35] 王科俊, 徐晶, 王磊, 等. 基于可拓遗传算法的机器人路径规划[J]. 哈尔滨工业大学学报, 2006, 38(7): 1135-1138.
- [36] 周明, 孙树栋, 彭炎午. 使用遗传算法规划移动机器人路径[J]. 西北工业大学学报, 1998, 16(4): 580-583.
- [37] 林磊, 王晓龙, 刘家锋. 基于遗传算法的手写体汉字识别系统优化方法的研究[J]. 计算机研究与发展, 2001, 38(6): 658-661.
- [38] 赵应丁, 刘金刚. 基于遗传算法的指纹图像二值化算法研究[J]. 计算机工程, 2006, 32(7): 169-171.
- [39] 王建锋, 吴庆标. 分层遗传算法实现图像边缘检测[J]. 计算机工程与应用, 2006, 42(14): 95-96, 151.
- [40] 王小平, 曹立明, 施鸿宝. 基于遗传算法的人工生命演示系统设计[J]. 同济大学学报: 自然科学版, 2003, 31(2): 224-228.
- [41] 查志琴, 高波, 郑成增. 遗传编程实现的研究[J]. 计算机应用, 2003, 23(7): 137-139.
- [42] KOZA J R. Genetic programming: on the programming of computers by means of natural selection[M]. Cambridge: MIT Press, 1992.
- [43] KOZA J R. Genetic programming II: automatic discovery of reusable programs[M]. Cambridge: MIT Press, 1994.
- [44] KOZA J R, BENNETT III F H, ANDRE D, *et al.* Four problems for which a computer program evolved by genetic programming is competitive with human performance[C] // Proc of IEEE International Conference on Evolutionary Computation. 1996: 1-10.
- [45] 文绍纯, 罗飞, 付连续. 遗传算法在人工神经网络中的应用综述[J]. 计算技术与自动化, 2001, 20(2): 1-5.
- [46] 承向军, 贺振欢, 杨肇夏. 基于遗传算法的交通信号机器学习控制方法[J]. 系统工程理论与实践, 2004, 24(8): 130-135.
- [47] 贾兆红, 倪志伟. 改进型遗传算法及其在数据挖掘中的应用[J]. 计算机应用, 2002, 22(9): 31-33.
- [48] 王静莲, 刘弘, 李少辉. 基于决策树的遗传算法在数据挖掘领域的应用[J]. 计算机工程与应用, 2005, 41(28): 153-155.
- [49] 张细政, 邢立宁, 伍栖. 基于遗传算法的数据挖掘方法及应用[J]. 哈尔滨工程大学学报, 2006, 27(7): 384-388.
- [50] SUNIL C. Design and implementation of a genetic based algorithm for data mining[C] // Proc of the 26th International Conference on Very Large Databases. San Francisco: Morgan Kaufmann Publishers, 2000: 33-42.

(上接第 2892 页)

- [13] 孙宏, 杨义先. 无线局域网协议 802.11 安全性分析[J]. 电子学报, 2003, 31(7): 1098-1100.
- [14] SMYTH N, McLOONE M, McANNY J V. WLAN security processor [J]. Circuits and Systems I: Regular Papers, IEEE Transactions, 2006, 53(7): 1506-1520.
- [15] ABOBA B, SIMON D. RFC 2716, PPP EAP TLS authentication protocol[S]. 1999.
- [16] DIERKS T, ALLEN C. RFC 2246, The TLS protocol version 1.0 [S]. 1999.
- [17] Cisco Systems. EAP-TLS deployment guide for wireless LAN networks [EB/OL]. (2004). http://www.cisco.com/en/us/tech/tk722/tk809/technologies-white-paper09186_a0080092566.shtml.
- [18] FUNK P, BLAKE-WILSON S. EAP tunneled TLS authentication protocol (EAP-TTLS) [S]. 2004.
- [19] PALEKEE A, SIMON D, JOSEFSSON S, *et al.* Protected EAP protocol (PEAP) [S]. 2004.
- [20] RIVEST R. RFC 1321, The MD5 message-digest algorithm[S]. 1992.
- [21] Cisco Systems. Lightweight extensible authentication protocol (LEAP) [S].
- [22] WU T. RFC 2945, The SRP authentication & key exchange system [S]. 2000.
- [23] BERSAI F, TSCHOFENIG H. RFC 4764, The EAP-PSK protocol: a pre-shared key extensible authentication protocol (EAP) method[S]. 2007.
- [24] CAM-WINGET N, MCGREW D, SALOWEY J, *et al.* RFC 4851, The flexible authentication via secure tunneling extensible authentication protocol method (EAP-FAST) [S]. 2007.
- [25] GRAYSON M, SALOWEY J. EAP SIM authentication: draft-haverinen-ppext-eap-sim-10.txt[R]. 2003.
- [26] ARKKO J, HAVERINEN H. RFC 4187, Extensible authentication protocol method for 3rd generation authentication and key agreement (EAP-AKA) [S]. 2006.
- [27] LIBERMAN D A. RSA security SecurID EAP[S]. 1999.