

应用微粒群算法确定有机化合物分子式

陈晓东¹, 张玉敏², 徐跃¹

CHEN Xiao-dong¹, ZHANG Yu-min², XU yue¹

1. 吉林大学 测试科学实验中心, 长春 130023

2. 吉林大学 化学学院, 长春 130023

1. Experimental Center of Testing Science, Jilin University, Changchun 130023, China

2. College of Chemistry, Jilin University, Changchun 130023, China

E-mail: cxd@jlu.edu.cn

CHEN Xiao-dong, ZHANG Yu-min, XU yue. Determining molecular formulas of organic compounds by Particle Swarm Optimization algorithm. *Computer Engineering and Applications*, 2009, 45(27): 246-248.

Abstract: To avoid the premature problem and the slow convergence of particle swarm optimization algorithm (PSO), an improved particle swarm optimization algorithm (IPSO) is presented to be used for determining molecular formulas of organic compounds. On the basis of integer programming and the PSO with contraction factor, the IPSO with mutation probability is proposed to get a good population diversity and to avoid PSO getting into local best result. The algorithm applied to determine molecular formulas of organic compounds is much better than those of PSO and PSO_HPO.

Key words: Particle Swarm Optimization algorithm; integer programming; mutation probability; mass fraction; molecular formulas

摘要: 针对基本微粒群优化算法 (PSO) 存在容易陷入局部最优和收敛速度慢的缺点, 在整数空间使用带收缩因子的微粒群优化算法基础上, 提出了一种带变异概率的微粒群优化算法 (IPSO), 用于提高微粒群的多样性, 避免算法陷入局部最优解。实验证明, 改进后的微粒群优化算法在防止早熟和加快收敛方面优于基本 PSO 算法和基本 PSO 算法加一半微粒随机初始化算法 (PSO_HPO 算法)。IPSO 算法应用到确定有机化合物分子式时, 取得了很好的效果。

关键词: 微粒群优化算法; 整数规划; 变异概率; 质量分数; 分子式

DOI: 10.3778/j.issn.1002-8331.2009.27.074 文章编号: 1002-8331(2009)27-0246-03 文献标识码: A 中图分类号: TP18

1 引言

表征一种有机化合物需要进行多种仪器分析来定量和定性该化合物。定量分析可使用元素分析仪测定有机化合物中碳、氢、氮、氧、硫等五种元素质量分数。对一些含有卤素、磷等非金属元素的有机化合物和金属有机化合物的定量分析, 还需借助化学分析法和其他分析仪器, 如用等离子发射光谱仪等, 确定其质量分数。定性分析可使用质谱仪确定有机化合物的分子量, 可使用红外光谱仪、C 谱核磁共振仪等对有机化合物进行结构分析, 最后确定该有机化合物的分子式。利用提供的方法, 可在一定程度上简化这一过程。利用改进的微粒群优化算法, 在已知有机化合物分子量和元素种类, 并通过元素分析测定其中某一元素的质量分数的情况下, 计算出该化合物的分子式。因此, 该算法在有机化合物的元素分析和有机化合物结构分析方面具有很好的应用前景。

2 微粒群优化算法

微粒群优化 (Particle Swarm Optimization, PSO) 算法是由 Kennedy 和 Eberhart 在 1995 年的 IEEE 国际神经网络学术会议上正式提出的一种基于群智能理论的优化算法^[1-2]。该算法通

过群体中微粒间的合作与竞争产生的群体智能指导优化搜索, 它既保留了传统优化算法利用群体智能的优点, 同时又具有操作模型简单, 容易理解, 易于实现的特点, 因此在许多优化问题中得到成功应用^[3-6]。

Shi 和 Eberhart 在 PSO 算法的速度项中引入了惯性权重 ω ^[6], 该改进算法一般称为基本 PSO 算法。在 D 维目标搜索空间中, 有 m 个微粒组成一个微粒群, 其中第 i 个微粒表示为 $X_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{id})$, $i=1, 2, 3, \dots, m$, 它以速度 V_i 在飞行中所经历的最好位置, 就是微粒 i 本身的最优位置, 记做 P_{id} (局部最优位置)。整个群体所经过的最优位置就是整个群体目前找到的最好位置, 记做 P_{gd} (全局最优位置)。每个微粒通过上述两个极值不断更新自己的位置, 从而产生新一代群体。实际操作中通过优化问题所决定的适应度函数来评价粒子位置的优劣程度。微粒 i 的速度用 $V_i = (v_{i1}, v_{i2}, v_{i3}, \dots, v_{id})$, $i=1, 2, 3, \dots, m$ 表示。对每一代, 其第 d 维 ($1 \leq d \leq D$) 按以下方程进化:

$$V_{id}(t+1) = \omega \times V_{id}(t) + c_1 \times r_1 \times (P_{id}(t) - X_{id}(t)) + c_2 \times r_2 \times (P_{gd}(t) - X_{id}(t)) \quad (1)$$

$$X_{id}(t+1) = X_{id}(t) + V_{id}(t+1) \quad (2)$$

式中, c_1, c_2 为常数, 称为正的加速因子; r_1 和 r_2 是 $[0, 1]$ 上均匀分

布的随机数; ω 为惯性权重, 用以确定该粒子先前速度的保留情况; P_{id} 是第 i 个微粒所经历的最好位置; P_{gd} 是所有微粒经历过的最佳位置。搜索被限制在最大、最小位置和最大、最小速度之间进行。

为了改善算法的收敛性, 文献[7]提出一种带收缩因子的 PSO 算法。其微粒 i 的位置按式(2)进化, 速度的迭代方程为:

$$V_{id}(t+1) = \chi \times (\omega \times V_{id}(t) + c_1 \times r_1 \times (P_{id}(t) - X_{id}(t)) + c_2 \times r_2 \times (P_{gd}(t) - X_{id}(t))) \quad (3)$$

式中 χ 是收缩因子, $\chi = 2 / (2 - \phi - \sqrt{\phi^2 - 4\phi})$, $\phi = c_1 + c_2$, $\phi > 4$, 不必将搜索限制在最大最小速度之间。有实验表明, 带收缩因子的 PSO 算法比不带收缩因子的 PSO 算法具有更好的性能^[8]。

3 带变异概率的微粒群优化算法

对式(1)的研究发现, 随着进化的进行, 如果微粒群局部最优位置 P_{id} 和全局最优位置 P_{gd} 保持较长时间不变, 微粒速度将沿着 $\omega \times V_{id}(t)$ 的方向变化, 微粒具有很强的“趋同性”, 此时算法可能陷入局部最优。为了解决这一问题, 曾有文献提出, 对适应度较低的一半微粒的位置和速度不按照式(1)和式(2)进行更新, 而是再一次随机初始化, 以提高微粒的多样性, 扩大搜索范围, 避免算法陷入局部最优。但是由于是随机对微粒的位置和速度初始化, 就不能保证每次计算均能得到最优解。

为了解决这一问题, 借鉴了遗传算法的变异算子方法, 该文设计了在整数空间和带收缩因子的 PSO 算法的基础上, 对适应度较低的一半微粒的速度和位置按逐步递减的变异概率进行变异的微粒群优化算法(Improved Particle Swarm Optimization, 以下简称 IPSO)。具体做法是, 对适应度较低的一半微粒的速度和位置不按照式(1)和式(2)进行更新, 而是按变异概率 P_m 进行变异。这样可以提高微粒的多样性, 扩大搜索空间, 避免算法陷入局部最优。该文设计与进化代数相关的变异概率 P_m , $P_m = 1 / (1 + 1.2 \times \ln(\text{gen}))$, 式中 gen 为目前的进化代数, 在进化初期, 变异的范围相对较大, 随着种群的进化, 变异的范围越来越小, 正是这种调节作用极大地加快了算法的收敛速度。图 1 是变异概率和进化代数的关系图。

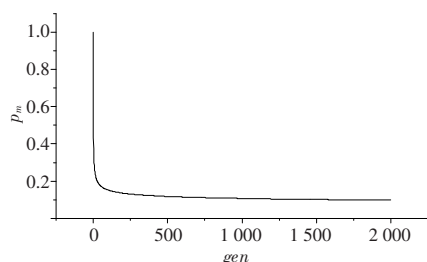


图 1 变异概率和进化代数的关系图

4 分子式的确定

确定有机化合物分子式的问题可归结为整数规划问题, 描述为: $\min f(x)$, $x \in S \subseteq Z^n$, 其中 Z^n 为 n 维整数空间, S 为一整数集, 可以为无界集。文献[9]提供了一种基于整数规划的 PSO 算法, 将进化限制在整数空间, 减少了许多不必要的非可行解域的搜索, 提高了计算效率。

确定有机化合物分子式的 IPSO 算法过程如下:

步骤 1 初始化种群。确定群体规模为 2 000, 代表碳、氢、氮、氧、硫五种元素的微粒位置范围为 $[0, 100)$, 以随机的方式给出每一个微粒的初始位置和速度。

步骤 2 评价。按式(4)计算每一个微粒的适应度函数值, 并和最好的适应度值相比较, 保存每一个微粒的最好适应度值和群体最好适应度值。该算法将接近目标分子量程度的函数作为适应度函数, 具体定义如下:

$$F(\omega) = \min((\sum(\omega_c, \omega_h, \omega_n, \omega_o, \omega_s) - M)) \quad (4)$$

其中, M 为目标分子量, F 为适应度函数, $\omega_c, \omega_h, \omega_n, \omega_o, \omega_s$ 分别为分子式中碳、氢、氮、氧、硫元素的质量分数。

步骤 3 找出每一个微粒到目前为止搜索过程中的最好位置 P_{id} 。

步骤 4 找出所有微粒到目前为止所搜索到的整体最好位置 P_{gd} 。

步骤 5 在整数空间依据式(2)与式(3)更新每一个微粒的速度与位置, 该算法中 $c_1 = 2.8, c_2 = 1.3, \omega = 1$ 。

步骤 6 对整个种群按微粒最优适应度进行排序, 对适应度较低的一半微粒的位置和速度不按照式(2)和式(3)进行更新, 而是按变异概率 P_m 进行变异。

步骤 7 返回步骤 2 继续执行, 直到得到一个令人满意的结果或符合终止条件为止。

根据该文的 IPSO 算法, 采用 Matlab 7.0 语言编制求取有机化合物分子式的程序。当给出某一含有 C、H、N、O、S 五种元素有机化合物的分子量为 240.291 8, 并已知其中 C 的质量分数为 29.99%, C、H、N、O、S 每种元素的位置范围为 $[0, 100)$ 时, 利用该程序可得到分子式 $C_6H_{12}N_2O_4S_2$ (胱氨酸标准物质)。

5 算法收敛性讨论

分以下三种情况讨论用微粒群算法确定有机化合物分子式的收敛性: (1) 使用基本 PSO 算法确定有机化合物的分子式。(2) 使用基本 PSO 加一半微粒随机初始化算法(以下简称 PSO_HPO 算法)^[10]确定有机化合物的分子式。(3) 使用 IPSO 算法确定有机化合物的分子式。目标分子式为 $C_6H_{12}N_2O_4S_2$, 代表碳、氢、氮、氧、硫五种元素的微粒位置范围为 $[0, 100)$, 运行次数为 100 次。比较结果详见表 1。

表 1 基本 PSO、PSO_HPSO、IPSO 三种算法的收敛性比较

算法名称	参数	最好适应度的平均值	收敛到全局最优解的次数	平均收敛代数
基本 PSO	$c_1 = c_2 = 2, V_{\max} = 10, \omega = 0.9 - (\text{gen}/\text{maxgen}) \times 0.5$	0.794 8	0	不收敛
PSO_HPO	同基本 PSO	0.416 4	92	247
IPSO	$c_1 = 2.8, c_2 = 1.3, \omega = 1$	0	100	26

注: 表中 gen 表示当前的进化代数, maxgen 表示最大进化代数, 为 2 000。

从表 1 可以看出, IPSO 算法无论从收敛速度还是从避免陷入局部最优方面都好于基本 PSO 算法和 PSO_HPO 算法。

6 结论和推广应用

(1) 提出了一种在整数空间使用带收缩因子和变异概率的微粒群优化算法确定有机化合物分子式的方法。对适应度较低的一半微粒的位置和速度不按照式(2)和式(3)进行更新, 而是按变异概率 P_m 进行变异。这样可以提高微粒的多样性, 扩大搜索空间, 避免算法陷入局部最优。实验证明, 该算法在防止早熟和加快收敛方面优于基本 PSO 算法和 PSO_HPO 算法。IPSO 算法在确定有机化合物分子式的应用中取得很好效果。

(2) 在已知有机化合物分子量及元素种类的情况下, 通过

元素分析测定其中某一元素的质量分数,利用该文的IPSO算法,可以迅速确定该有机化合物的分子式。

(3)在仅仅知道有机化合物分子量和元素种类的情况下,利用该算法也可算出该化合物的分子式,但得到的分子式可能不唯一,有可能得到多个同分异构体分子式。

(4)推广该文的IPSO算法,在已知任意化合物分子量和元素种类以及该化合物中任何一种元素(最好为含量高的)质量分数的情况下,对任意化合物的分子式进行确定。

参考文献:

- [1] Kennedy J, Eberhart R C. Particle Swarm Optimization[C]//IEEE International Conference on Neural Networks: Perth, Piscataway, NJ, Australia; IEEE Service Center, 1995, IV: 1942-1948.
- [2] Eberhart R C, Kennedy J A. A new optimizer using particle swarm theory[C]//Proceedings of the Sixth International Symposium on Micro Machine and Human Science. New York, USA: IEEE, 1995: 39-43.
- [3] Parsopoulos K E, Vrahatis M N. Particle swarm optimization method for constrained optimization problems[J]. Intelligent Technologies Theory and Applications: New Trends in Intelligent Technologies, 2002: 214-220.
- [4] Eberhart R C, Hu X. Human tremor analysis using particle swarm

optimization[C]//Proceedings of the IEEE Congress on Evolutionary Computation (CEC 1999), Washington D C, 1999: 1927-1930.

- [5] Yoshida H, Kawata K, Fukuyama Y, et al. A particle swarm optimization for reactive power and voltage control considering voltage security assessment[J]. IEEE Transactions on Power Systems, 2000: 1232-1239.
- [6] Shi Y, Eberhart R C. A modified particle swarm optimizer[C]//IEEE Int'l Conf on Evolutionary Computation. Anchorage, Alaska, 1998: 69-73.
- [7] Clerc M. The swarm and the queen: Towards a deterministic and adaptive particle swarm optimization[C]//Proceedings of the 1999 Congress on Evolutionary Computation. Piscataway, NJ, USA: IEEE, 1999: 1927-1930.
- [8] Eberhart R C, Shi Y. Comparing inertia weights and constriction factors in particle swarm optimization[C]//Proceedings of the 2000 Congress on Evolutionary Computation. Piscataway, NJ, USA: IEEE, 2000: 84-88.
- [9] 谭瑛, 高慧敏, 曾建潮. 求解整数规划问题的微粒群算法[J]. 系统工程理论与实践, 2004(5): 126-129.
- [10] 张利彪, 周春光, 刘小华, 等. 粒子群算法在求解优化问题中的应用[J]. 吉林大学学报: 信息科学版, 2005, 23(4): 385-389.

(上接 205 页)

素在模拟中没有很好地反应等原因,具有一定的局限性。多智能体系统(Multi-Agent System)建模是一种优秀的仿真理论和方法,把元胞、元胞空间抽象为智能体(Agent),可以提高元胞选择径路时的智能性,已成为研究人群疏散的研究趋势。在下一步工作中,将对元胞的移动规则进行更微观地分析研究,利用多智能体相关理论与元胞自动机理论相结合,来不断修正和完善模型。

参考文献:

- [1] 王理达. 地铁站人群疏散行为仿真研究[D]. 北京: 北京交通大学, 2006: 3-5.
- [2] 王驰. 某地铁站火灾情况下人员安全疏散研究[D]. 北京: 北京交通大学, 2007: 3-4.
- [3] 胡忠日. 安全疏散研究的国内外动态和发展趋势[J]. 消防科学与技术, 2001(6): 7-8.

术, 2001(6): 7-8.

- [4] 蒋济同, 李华军. 火灾时人员安全疏散可靠性评估[J]. 火灾科学, 1997, 6(1): 7-11.
- [5] 尹宏伟, 张乐年, 蔡守允. 水利模型实验控制系统[J]. 工业控制计算机, 2006, 19(9): 17-18.
- [6] 王辉, 钱锋. 一种基于距离行为模型的改进微粒群算法[J]. 计算机工程与应用, 2007, 43(30): 30-32.
- [7] 方伟峰, 杨立中, 黄锐. 基于元胞自动机的多主体人员行为模型及其在性能化设计中的应用[J]. 中国工程科学, 2003, 5(3): 67-71.
- [8] Bradley G E. A proposed mathematical model for computer prediction of crowd movements and their associated risks[C]//Smith R A, Dickie J F. Engineering for Crowd Safety. [S.l.]: Elsevier Science Publisher, 1993.
- [9] Schadschneider A. Pedestrian and evacuation dynamics [M]. Berlin: Springer, 2002.

(上接 239 页)

参考文献:

- [1] Spitzer V M, Whitlock D G. The visible Human dataset: the anatomical platform for human simulation[J]. Anat Rec, 1998, 253(2): 49-57.
- [2] Kawata Y, Ninin, Ohmatsu H, et al. A deformable surface model based on boundary and region information for pulmonary module segmentation from 3D thoracic CT images[J]. IEICE Transactions on Information and Systems, 2003, E86-D(9): 1921-1930.
- [3] The Dicom Standard[S/OL]. <http://medical.nema.org/nema/medical>.
- [4] 陈贤志. 医学数字图像通信(DICOM)标准介绍[J]. 中国医学影像学杂志, 1995(4): 204-207.
- [5] 孙亦博. 基于CT数据的三维重构相关技术的研究与应用[D]. 西安: 西北大学, 2007.
- [6] Parker J R. Gray level thresholding in badly illustrated images[J]. IEEE Trans on Pattern Analysis and Machine Intelligence,

1991, 13(8): 813-819.

- [7] Dizenzo S, Cinque L, Levialdi S. Image thresholding using fuzzy entropies[J]. IEEE Trans on System, Man, and Cybernetics, B, 1998, 28(1): 15-23.
- [8] 夏良正. 数字图像处理[M]. 南京: 东南大学出版社, 1999.
- [9] 章毓晋. 图像分割[M]. 北京: 科学出版社, 2001.
- [10] 徐升, 唐庆玉. X-CT体数据的人体骨骼分割算法[J]. 北京生物医学工程, 2002, 21(3): 175-178.
- [11] 田捷, 包尚联, 周明全. 医学影响处理与分析[M]. 北京: 电子工业出版社, 2003.
- [12] Gonzalez R C, Woods R E. 数字图像处理[M]. 北京: 电子工业出版社, 2006.
- [13] Haralic R M, Sternberg S R. Image analysis using mathematical morphology[J]. IEEE Trans on PAMI, 1987, 9(4): 532-536.
- [14] Senger S. Visualizing and segmenting large volumetric data sets[J]. IEEE Computer Graphics and Applications, 1999(5/6): 32-37.