文章编号:1671-4229(2023)06-0077-05

基于因素空间的模糊多种群粒子群算法

钟育彬, 范书衡
(广州大学 数学与信息科学学院, 广东 广州 510006)

摘 要: 粒子群算法(Particle Swarm Optimizer, PSO)在处理较大数据的寻优问题时存在运行时间过长,易陷入局部最优的问题。文章基于传统粒子群算法,利用基于因素空间的特征选取法,优化特征数据组合,减少算法运行时间,利用多种群粒子群优化模型思想和模糊数学思想相结合提出基于因素空间的模糊多种群粒子群算法,增强了算法全局寻优能力并加快收敛速度,实现了算法的总体最优,改善了传统粒子群优化算法容易收敛到局部最优的缺陷。研究结果表明,与传统粒子群算法、动态多种群粒子群算法、基于密度峰值的多种群粒子群算法相比,文章所提出的基于因素空间的模糊多种群粒子群算法具有更优的寻找全局最优解能力和收敛速度。

关键词: 因素空间; 多种群粒子群算法; 优化问题
中图分类号: O29 **文献标志码**: A

Fuzzy multi-population particle swarm optimization based on factor space

ZHONG Yu-bin, FAN Shu-heng
(School of Mathematics and Information Science, Guangzhou University, Guangzhou 510006, China)

Abstract: Particle Swarm Optimizer (PSO) has the problems of long running time and easy to fall into local optimality when dealing with the optimization of large data. In this paper, based on the traditional particle swarm optimization algorithm, the feature selection method based on factor space is used to optimize the combination of feature data and reduce the running time of the algorithm. The fuzzy multi-population particle swarm optimization algorithm based on factor space is proposed by combining the idea of multi-population particle swarm optimization model with the idea of fuzzy mathematics, which enhances the global optimization ability of the algorithm and speeds up the convergence rate, thus achieving the overall optimization of the algorithm. The traditional particle swarm optimization algorithm is easy to converge to the local optimal defect. The results show that compared with the traditional particle swarm optimization algorithm, dynamic multi-population particle swarm optimization algorithm and multi-population particle swarm optimization algorithm based on density peak, the fuzzy multi-population particle swarm optimization algorithm based on factor space proposed in this paper has a better ability to find the global optimal solution and convergence speed.

Key words: factor space; multi-population particle swarm optimization; problem of optimization

随着机器学习的快速发展,人工智能已经运用到多种不同领域的现实问题中。机器学习同样能够在癌症候选药物领域学习分析出不同分子描述符及其取值范围针对不同疾病相关靶标的治疗效果[1-3]。在机器学习目标约束函数寻优领域中,有多种常用寻优算法[4],如粒子群优化算法、蚁群算法、遗传算法、模拟退火算法等。其中,粒子群算法(Particle Swarm Optimization, PSO)[5],是一种基于鸟群捕食行为研究的进化计算方法,PSO由Kennedy和Eberhart博士在1995年提出。PSO的

收稿日期: 2023-04-07; **修回日期**: 2023-05-31
作者简介: 钟育彬(1965—),男,教授. E-mail:zhong-yb@163.com
引文格式: 钟育彬, 范书衡. 基于因素空间的模糊多种群粒子群算法[J]. 广州大学学报(自然科学版),2023,22(6):77-81.

基本概念是鸟群在随机搜寻一片区域中唯一的食物,虽然所有鸟并不知道食物的位置,但它们能够感知到当前位置与食物的距离。此时,最有效的搜寻策略就是在离食物最近的鸟的附近区域搜寻。

粒子群算法的缺点在于容易陷入局部最优,导致收敛精度较低或者不易收敛。本文在基于因素空间[6-9]的特征选取法[10]选取特征数据的基础上,利用多种群粒子群[11-14]优化模型思想和模糊数学思想相结合提出基于因素空间的模糊多种群粒子群算法。首先,利用基于因素空间的特征选取法识别有效的且区分度较大的特征变量,降低粒子维度,得到加快粒子群算法运算的效果;其次,通过保持记录最优的5个全局最优解,集合多种群粒子群模型、多种群并行进化与模糊综合评价思想,赋予粒子新的收敛方向,从而在增强算法搜索全局最优解能力的同时,加快收敛速度。

1 传统粒子群算法计算时间复杂度分析

目标约束函数寻优算法所需的运算时间不仅与数据量的大小有关,也和目标函数与约束条件的复杂程度有关。对于简单的目标函数与约束条件,传统粒子群优化算法在计算每个粒子所在位置与目标距离的计算时间较短,因此,通常能够较快地完成整个寻优运算,并得到良好的结果。反之,当目标函数与约束条件较为复杂时,计算每个粒子当前位置与目标距离所需的运算量增加,以致于完成整个寻优算法所需的运算时间大大增加。

本文中乳腺癌候选药物含有种类较多的分子描述符,通过随机森林和XGBoost分类模型预测不同分子描述符组合对不同靶标的作用效果,以多个靶标效果预测结果作为目标函数,再利用粒子群算法对分子描述符的用量做最大优化。其中,由于分子描述符种类较多,数据量较大,传统粒子群优化算法的运算量较大,所需运算时间较长。本部分对传统粒子群优化算法进行时间复杂度分析,假设粒子的数量不变,数目为n,最大迭代次数为m,每一个粒子经历一次迭代的时间为t,那么,对于传统粒子群优化算法进行寻优所需要的总的计算时间为

$$T = n \times m \times t。$$

假设设定模型参数中粒子的数量为100个,迭代次数为100次,计算一次粒子当前位置的适应度函数值时间为0.1 s,在每次迭代中,每个粒子需要进行6次适应性函数值计算,则传统粒子群优化算法寻优过程中在计算适应性函数所需要的时间为100 min。

为了在加快粒子群算法寻优收敛速度的同时增强算法寻找全局最优解的能力,本文提出基于因素空间的模糊多种群粒子群算法。

2 基于因素空间的模糊多种群粒子群算法

改进粒子群优化算法收敛速度的关键在于找到最好的粒子位置,根据最好的粒子位置为其余粒子提供更清晰准确的寻找方向。本文运用因素空间理论中成熟的模糊综合评价体系,能够在众多位置较优粒子中选取或结合出位置更优的粒子,以此加快粒子群优化算法的收敛速度。另外,粒子群优化算法具有易陷入局部最优解的缺点,改进这个缺点的关键在于如何让新一代的粒子跳出当前局部最优解的领域,本文利用多种群粒子群思想由多个种群进行同时并行进化,增强算法寻找全局最优解的能力。

模糊综合评价体系的方法运用了模糊数学的理论与思想,该方法是以模糊运算和最大隶属度原则为主,对模糊因素进行解析化和定量化。将模糊综合评价体系与粒子群算法相结合,利用模糊综合评价加强粒子对当前位置优劣的判断,也使得结果更具有科学性,更易于分析解释多目标趋向。图1为模糊综合评价流程图。

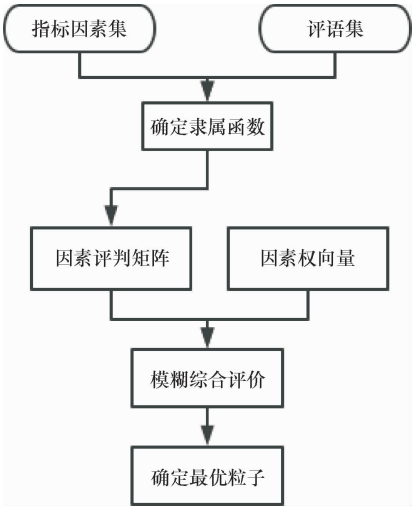

图1 模糊综合评价流程图

Fig. 1 Flow chart of fuzzy comprehensive evaluation

设定综合评价因素矩阵:

$$F=[f_1,f_2,f_3,\cdots,f_n],$$

综合评价因素矩阵 F 为影响判断粒子当前位置优劣的若干因素所组成的矩阵。

建立相应评语矩阵:

$$V=[v_1,v_2,v_3,\cdots,v_m],$$

评语矩阵是对评价对象做出相应各评价因素做出的各种评价。如确定为 5 个级别,即优秀、良好、较好、一般和较差。

因素权重矩阵:

$$I=[i_1,i_2,i_3,\cdots,i_n]$$

评价因素矩阵中的各因素重要程度不同,因此,在评价过程中,对每一个因素 f_i 的权重不同,通过对每个评价因素赋予权重,获得因素权重矩阵 I。

建立因素评判矩阵,用 r_{ij} 表示第 i 个因素获得第 j 个评价的概率,即 i 对 j 的隶属度,每个因素可以得到一个单因素评判矩阵 R_i。

$$R_i=[r_{i1},r_{i2},r_{i3},\cdots,r_{im}]。$$

由 n 个因素所得 $n\times m$ 的因素评判矩阵 R,记为

$$R=\begin{bmatrix} r_{11}, & r_{12}, & \cdots, & r_{1m} \\ r_{21}, & r_{22}, & \cdots, & r_{2m} \\ \vdots, & \vdots, & \cdots, & \vdots \\ r_{n1}, & r_{n2}, & \cdots, & r_{nm} \end{bmatrix}。$$

根据建立好的因素权重矩阵 I 与因素评判矩阵 R 得到模糊综合评价,记为

$$M=I\times R。$$

根据 M 的结果可以得到最好的粒子位置,在粒子群优化算法框架中提供更优秀的寻找方向。通过模糊综合评价体系与粒子群算法的结合大幅度加快了粒子的收敛速度,此时,利用多种群粒子群思想,令多个种群同时进化,每隔 5 代进行信息交互,从而达到增强算法,寻找全局最优解能力的效果。

步骤 1:建立模糊综合评价体系。

根据复杂适应性函数多目标的影响因素及重要程度等建立模糊综合评价体系,将模糊综合评价体系加入到多种群粒子群算法框架中更新粒子位置部分。

步骤 2:选取最优粒子。

通过适应性函数计算依次找出若干个全局最优解,确定因素集 F 等,获得模糊综合评价 M 选择最优粒子以及 i 个次优粒子,赋予权重更新新一代粒子位置。可以通过希望获取更快的收敛速度或者更强的寻找全局最优解能力两个思路确定选取粒子的数量 i。

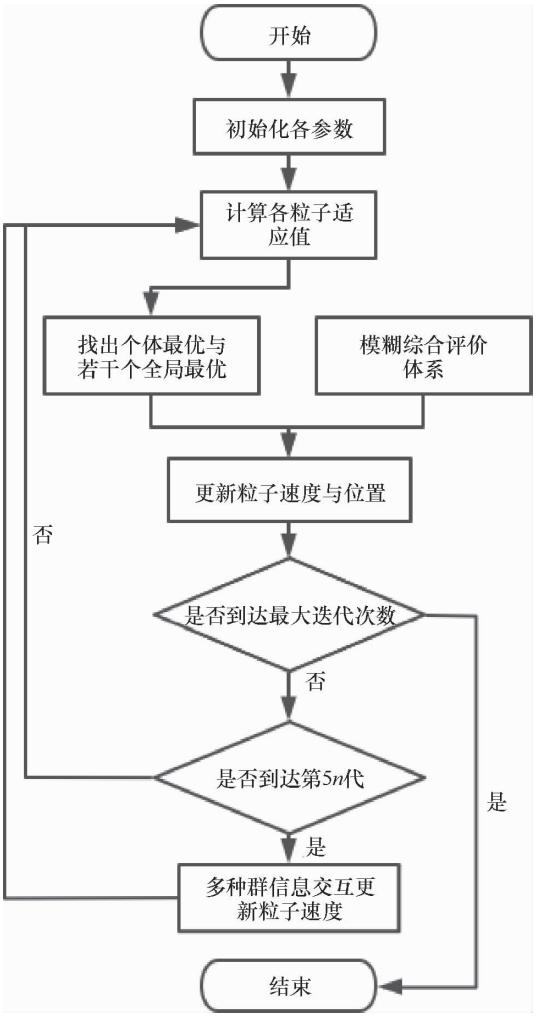

图 2　基于因素空间的模糊多种群粒子群算法流程图

Fig. 2　Flow chart of fuzzy multi-population particle swarm optimization based on factor space

步骤 3:信息交互。

由于粒子群通过个体最优粒子与全局最优粒子信息在寻优空间中快速收敛,具有易陷入局部最优解的问题。在每隔 5 代进行多种群信息交互,以新的全局最优粒子提供收敛方向,增强粒子的全局寻优能力。

步骤 4:循环迭代。

重复步骤 2 ~ 3,达到算法预设迭代次数后,算法停止。

3 算法时间复杂度分析

当基于因素空间的模糊多种群粒子群算法的种群数量为 x,迭代次数设为 n,每个种群粒子数量为 m,计算粒子当前位置模糊综合评价所需计算量为 nm,每代模糊综合评价作比较的计算量为 m,多种群信息交互计算量为 $x/5$,故本文改进点的计算量为

$$T = 2nm + x/5。$$

本文加入的算法时间复杂度为 $o(n^2)$,那么对于传统粒子群算法,其适应性函数内除了目标函数与约束条件框架外,还有计算量大于上式的程序操作,可选用本文改进方法进行操作。

4 实例分析

本文以抗乳腺癌候选药物选取为例,数据源于 2021 年中国研究生数学建模竞赛 D 题“抗乳腺癌候选药物的优化建模”。本文对雌激素受体 α 亚型(Estrogen Receptors alpha, ERα)靶标生物活性值进行预测,通过多种回归模型评估,LightGBM 的预测效果优于其他回归模型。模型参数设置如表 1 所示。

表 1 LightGBM 参数取值表

Table 1 LightGBM parameter value table

参数名	参数取值
n_estimators	250.0
learning_rate	0.1
min_data_in_leaf	41.0
bagging_fraction	0.7
feature_fraction	0.6
max_depth	7.0
max_bin	25.0

此外,一个化合物想要成为候选药物,除了需要具备良好的生物活性外,还需要在人体内具备良好的药代动力学性质和安全性,合称为 ADMET(Absorption 吸收、Distribution 分布、Metabolism 代谢、Excretion 排泄、Toxicity 毒性)性质。本文通过训练多种分类预测模型,决定分别选用 XGBoost 和 RandomForest 两种分类预测模型对化合物的 5 种性质 Caco-2、CYP3A4、hERG、HOB、MN 进行分类预测。Caco-2 代表该化合物的小肠上皮细胞渗透性,CYP3A4 代表该化合物是否能够被代谢,hERG 代表该化合物是否具有心脏毒性,HOB 代表该化合物的口服生物利用度,MN 代表该化合物是否具有遗传毒性。5 个分类预测模型由网格搜索调参,参数如表 2 所示。

表 2 XGBoost 及 RandomForest 参数取值表

Table 2 XGBoost、RandomForest parameter value table

ADMET 性质	最佳二分类模型	n_estimators 参数取值	max_depth 参数取值
Caco-2	XGBoost	200	23
CYP3A4	XGBoost	140	10
hERG	XGBoost	180	14
HOB	RandomForest	180	14
MN	RandomForest	180	14

针对以上 6 个模型作为多目标复杂适应性函数,使用基于因素空间的模糊多种群粒子群算法进行寻优,参数设置为:迭代次数 100,种群大小 100,结果见图 3。

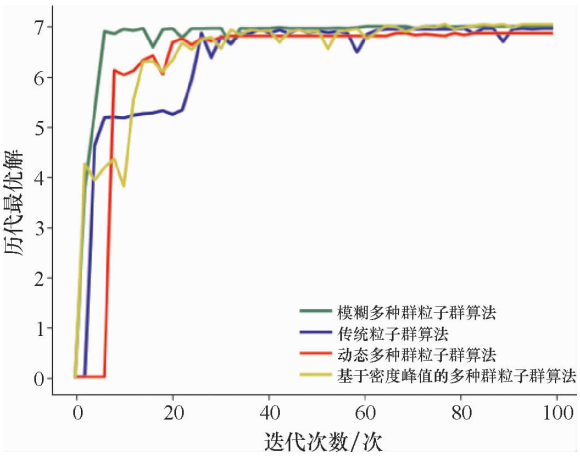

图 3 改进算法结果对比

Fig. 3 The results of the improved algorithm are compared

本文选用传统粒子群算法、动态多种群粒子群算法[15],以及基于密度峰值的多种群粒子群算法[16]作结果对比分析。由图 3、表 3 可知,基于因素空间的模糊多种群粒子群算法在确保获得最优解的同时,能够加快粒子收敛速度,比传统粒子群算法、动态多种群粒子群算法以及基于密度峰值的多种群粒子群算法更快地向最优解收敛。

表 3　实验结果对比表

Table 3　Comparison table of experimental results

模型名称	平均迭代次数	最优生物活性值	最优迭代时间
模糊多种群粒子群算法	19	6.957 6	179.285 3″
传统粒子群算法	28	6.894 7	196.548 6″
动态多种群粒子群算法	20	6.885 6	182.386 7″
基于密度峰值的多种群粒子群算法	21	6.955 8	185.354 8″

模糊多种群粒子群算法主要优势体现在每次迭代中,能更精准地找到当前最优粒子位置,减少最优迭代次数,从而减少算法寻求最优解所需时间。

5　结　论

本文提出的基于因素空间的模糊多种群粒子群算法以因素空间理论中模糊综合评价体系寻找最优粒子个体,加快收敛速度;结合多种群粒子群算法避免陷入局部最优。相比传统粒子群算法、动态多种群粒子群算法和基于密度峰值的多种群粒子群算法,基于因素空间的模糊多种群粒子群算法具有更优的寻找全局最优解能力与收敛速度。

本文改进算法所需时间复杂度不高,适用性强。

参考文献:

[1] Yujie Y, Xin L, Yi P, et al. Artificial intelligence in cancer target identification and drug discovery[J]. Signal Transduction and Targeted Therapy, 2022,7(6):24.

[2] Castelli M, Serapian S A, Marchetti F, et al. New perspectives in cancer drug development: Computational advances with an eye to design[J]. RSC Medicinal Chemistry, 2021, 12(9):1491-1502.

[3] Ghosh A, Fan Y, Li H Y, et al. Artificial intelligence and cancer drug development[J]. Recent Patents on Anti-Cancer Drug Discovery, 2022, 17(1):2-8.

[4] 王先超,韩波,张开银,等. 粒子群算法在求解数学建模最优化问题中的应用[J]. 阜阳师范学院学报(自然科学版),2016,33(2):117-121.

[5] Kennedy J. Particle swarm optimization[J]. Springer US, 1995(1). doi:10.1107/S11721-007-0002-0.

[6] 汪培庄. 因素空间与因素库[J]. 辽宁工程技术大学学报(自然科学版),2013,32(10):1297-1304.

[7] 汪培庄. 因素空间理论——机制主义人工智能理论的数学基础[J]. 智能系统学报,2018,13(1):37-54.

[8] 汪华东,汪培庄,郭嗣琮. 因素空间中改进的因素分析法[J]. 辽宁工程技术大学学报(自然科学版),2015,34(4):539-544.

[9] 汪培庄,周红军,何华灿,等. 因素表示的信息空间与广义概率逻辑[J]. 智能系统学报,2019,14(5):843-852.

[10] 刘玉铭. 因素空间与分类的特征抽取[J]. 北京师范大学学报(自然科学版),2000(2):172-177.

[11] 张天佳,杨永胜. 基于并行结构的多种群粒子群优化算法[J]. 传感器与微系统,2020,39(9):119-121.

[12] 向毅. 基于 Nelder-Mead 单纯形法的混合动态多种群粒子群优化算法的研究与应用[D]. 广州:广州大学,2013.

[13] Zhang W, Ran Y, Zhang G B, et al. Optimal allocation of product reliability using novel multi-population particle swarm optimization algorithm[J]. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2022, 236(9):4565-4576.

[14] Li G Q, Wang W L, Zhang W W, et al. Grid search based multi-population particle swarm optimization algorithm for multimodal multi-objective optimization[J]. Swarm and Evolutionary Computation, 2021,62, doi:10.1016/J.SWEVO.2021.100843.

[15] 于慧,王宇嘉,陈强,等. 基于多种群动态协同的多目标粒子群算法[J]. 电子科技,2019,32(10):28-33.

[16] 陶新民,郭文杰,李向可,等. 基于密度峰值的依维度重置多种群粒子群算法[J]. 软件学报,2023,34(4):1850-1869.

【责任编辑:陈　钢】