1. 引言

随着计算科学的不断发展，遗传算法（Genetic Algorithm, GA）作为一种基于生物进化理论的优化方法，已经在众多领域得到了广泛应用。自20世纪70年代由约翰·霍兰德（John Holland）提出以来，遗传算法逐步发展为解决复杂优化问题的强有力工具[6, 22]。其核心思想通过模拟自然选择、交叉与变异等生物学过程，结合适应度函数，能够在庞大的解空间中进行有效搜索，找到全局最优解。

遗传算法作为一种元启发式算法，其在优化问题中的应用已经涵盖了广泛的领域，如经济学、工程学、政治学、管理学等多个学科中，都能看到其应用的身影。遗传算法在解决诸如函数优化、机器学习、进化编程等问题时，展示出了显著的优势，特别是在面对复杂的、非线性、高维、约束条件较强的问题时，遗传算法展现出了无与伦比的灵活性和适应性。

遗传算法的基本流程主要包含[24]：

1. **选择操作**：选择操作基于个体的适应度进行筛选，选择适应度较高的个体作为父代，以保证优秀基因能够遗传到下一代。选择操作的方式多种多样，包括轮盘赌选择、锦标赛选择等，能够根据具体问题的需求进行调整和优化。
2. **交叉操作**：交叉操作通过交换父代染色体的部分基因产生新的后代。交叉操作的目的在于融合父母双方的优良基因，从而产生具有更好适应度的后代。交叉算子有很多种形式，常见的有单点交叉、双点交叉、均匀交叉等，每种交叉方式都有其特定的应用场景。
3. **变异操作**：变异操作则通过随机改变染色体上的某些基因来维持种群多样性，避免种群过早收敛到局部最优解。变异操作增强了遗传算法的探索能力，使得算法能够在全局范围内进行更广泛的搜索，避免陷入局部最优解的困境。

然而，尽管遗传算法取得了显著的应用成果，仍然存在一些挑战。例如，如何提高算法的收敛速度、避免早熟收敛、提高解的精度等问题，依然是当前遗传算法研究的热点之一。为了解决这些问题，研究人员对遗传算法进行了多方面的改进，并有效避免算法陷入局部最优解的困境。

1. 相关研究
   1. 编码策略

编码是将可行解从其解空间转换为遗传算法能够处理的搜索空间的方法。在编码方式的选择上，Holland提议采用二进制编码。HE Y 等人提出了一种用于多目标进程调度的二进制编码遗传算法，算法中通过对关键产品、任务和单位的识别，选择性地编码部分二进制变量[5]。YANG X等人[12]介绍了一种混沌Gray编码方法，解决了传统二进制编码遗传算法在数值优化时存在的精度问题。为了提高精度，陈辉等人[15]提出了一种将实数编码与量子位概率指导的交叉和混沌变异相结合的量子位编码方法，适用于大空间搜索问题。

尽管实数编码具有较高精度和更好的性能，但其主要适用于连续变量问题。LIN Y-C等人[9]还提出了混合染色体编码方式，作为一种有效的编码方法。梁旭等人[21]则将复数编码的思想应用于遗传算法中，通过实部和虚部表示变量，以此挖掘个体的多样性，减少局部收敛问题。郑朝晖等人[18]提出了一种动态相似度参数零件族编码方法，根据零件的工艺和相似性动态划分零件族，显著减少了编码长度和计算时间。

* 1. 遗传算子
     1. 交叉算子

金晶等人提出对选择出的父代个体进行有向交叉。根据父代个体的位置关系进行优化，确保子代个体落在靠近最优解的区域。设定父代个体，通过优化交叉点的位置，使子代个体更有可能朝最优解方向进化，确保了子代个体的适应度优于父代，避免盲目搜索，提高收敛速度[17]。

Deep等人提出一种新型的实数编码遗传算法交叉算子**（LX）**，使用拉普拉斯分布生成后代，与两种变异技术（MPTM和NUM）结合，形成了两种遗传算法：LX-MPTM和LX-NUM。拉普拉斯交叉算子在多个全局优化问题中表现优异，LX交叉算子在收敛速度和解的质量上均有显著提高[2]。

* + 1. 选择算子

乔家庆等人提出了引入个体差异概念的新选择算子，该算子在选择过程中同时考虑适应度和个体差异。这种基于差异的选择策略能够更全面地探索解空间，有效避免早熟收敛，尤其适用于多峰函数优化问题。实验表明，该方法具有更好的收敛性能和解的质量[20] 。

Pavez-Lazo等人提出了两个关键改进：确定性选择策略和环形交叉算子。确定性选择策略通过同时考虑适应度和个体间的遗传信息交换，避免了简单的优胜劣汰，有效增强了种群多样性。而环形交叉算子则创新性地将染色体构造为环状结构，克服了传统线性交叉在处理规划周期首尾时段时的信息交换障碍，从而更好地优化了发电机组的开停机方案[10]。

* + 1. 变异算子

陈峰等人提出了一种定向变异算子，用于解决遗传算法中的“欺骗问题”。通过引入特定的变异方向，增强了遗传算法跳出局部最优解的能力，从而提高了算法的全局搜索能力，实验表明定向变异算子能够显著提高算法的搜索效率和解的质量，尤其在多峰问题和高维问题中表现突出[14]。

Kusum等人提出了一种**幂变异**，应用于实数编码遗传算法。通过幂分布控制变异强度，使得算法能够根据问题的复杂性调整变异的幅度，从而提高全局优化能力。实验表明，幂变异的遗传算法能够更快速且准确地找到全局最优解，尤其在处理高维和多峰优化问题时表现更为优秀[3]。

* 1. 参数确定

遗传算法的参数主要包括：染色体位串长度l、群体规模n、交叉概率以及变异概率。为寻找最优参，产生了许多研究：模糊规则控制选择和变异率[8]，利用云模型云滴的随机性和倾向性特点，借助条件发生器产生交叉和变异率[16]。

* 1. 收敛性

彭茜等人基于高效遗传算法的电网需求侧调度优化问题，通过优化需求侧调度策略，提出了一种改进的遗传算法。通过多个电网系统的测试，验证了该高效遗传算法具有更好的收敛性能和解的精度，能在合理时间内找到最优调度方案[19]。

白晓明等人针对差分遗传算法中的收敛性问题进行了深入的研究。分析了其在不同优化问题中的收敛性表现，并提出了提升收敛效率的策略。实验结果表明，差分遗传算法在测试函数上的表现优于传统遗传算法和标准差分进化算法，证明了差分遗传算法在收敛性和优化效率方面的优势[13]。

1. 原理研究
   1. 算法步骤

遗传算法的数学模型可表示为一个七元组。

|  |  |
| --- | --- |
|  |  |

其中：C：编码空间，E：评价函数，P₀：初始种群，M：种群规模，Φ：遗传算子集合，Ψ：终止准则，T：最大代数

初始化种群：在解空间中随机生成M个个体作为初始种群，将问题的解空间映射到遗传空间。

|  |  |
| --- | --- |
|  |  |
|  |  |

其中为第个个体，表示第个基因位。

对每个个体计算其适应度值，用于评价个体优劣程度的量化指标：

|  |  |
| --- | --- |
|  |  |

其中为适应度函数，为目标函数。

选择操作：基于适应度的比例，个体i被选中的概率。

|  |  |
| --- | --- |
|  |  |

交叉操作：设交叉概率为Pc，对两个父代个体在位置k进行单点交叉。

|  |  |
| --- | --- |
|  |  |

其中为第*i*个子代，为第*i*个父代。

变异操作：设变异概率为，对个体X的第*j*位进行变异。

|  |  |
| --- | --- |
|  |  |

进化过程：种群进化的数学表达式。

|  |  |
| --- | --- |
|  |  |

终止条件：算法在满足以下条件之一时终止。

|  |  |
| --- | --- |
|  |  |

其中，为最大代数，为预设的精度要求。

* 1. 机理分析
     1. 杂交操作

杂交操作通过重组父代个体的基因来生成子代个体，从而探索更广阔的解空间。不同类型的杂交操作影响着遗传算法的表现和效率。

|  |  |
| --- | --- |
|  |  |

杂交操作通过结合父代个体的基因信息生成新解。从而，创造出具有较高适应度的后代，同时，不同的杂交方式有助于在解空间中不同的区域进行广泛搜索[11]。杂交操作能有效避免遗传算法在局部最优解附近陷入困境[24]。通过交叉产生新的解，遗传算法能够在不同的区域进行搜索，从而跳出可能的局部最优解。

* + 1. 隐性并行

遗传算法的“隐性并行”是其一大优势，体现在群体中多个个体的同时评估和操作上。在遗传算法中，解空间的探索通过多个个体在每一代中并行地进行选择、交叉和变异操作，使得算法能够同时在多个解空间区域内进行探索。

|  |  |
| --- | --- |
|  |  |

其中，是当前代模式H的个体数，是模式H的适应度，F是群体的平局适应度。

根据模式定理[1]和，群体并行评估不仅有效避免了"早熟收敛"现象，增强了全局搜索能力，还使其能够充分利用现代分布式计算平台，在处理大规模优化问题时表现出优异的可扩展性和计算效率[23]。

* + 1. 优胜劣汰的影响

|  |  |
| --- | --- |
|  |  |

其中，和分别为出生和死亡的个体数量。

优胜劣汰通过优先选择适应度较高的个体，确保优良基因得以保留并在下一代得以繁殖，加速了算法的收敛过程。通过在每一代中保留适应度较高的个体，能够有效地减少计算时间，增加高适应度模式个体浓度，更好地利用有效的模式。推动了遗传算法向全局最优解的收敛[4]。

1. 遗传算法应用
   1. 数据集：

托卡特加齐奥斯曼帕夏大学医学院的贫血诊断数据集[7]，包含了2013年至2018年间15,300名患者的完整血液检查结果，保留了全部24个特征，包含五种类型：64%的无贫血患者、27%的铁缺乏性贫血、7%的HGB贫血、1%的B12缺乏性贫血和1%的叶酸缺乏性贫血。

* 1. 实验配置：

1. 遗传算法

运用Pygad包，迭代次数：100，父代：4，种群大小：50。

1. 模拟退火

运用Simanneal包，初始温度：25，最终温度：0.001，总步数：1000。

1. 粒子群

运用PYSwarms包，粒子：30，认知参数：0.5，社会参数：0.3，怪形权重：0.9。

1. 爬山算法

迭代次数：1000。

* 1. 实验结果：

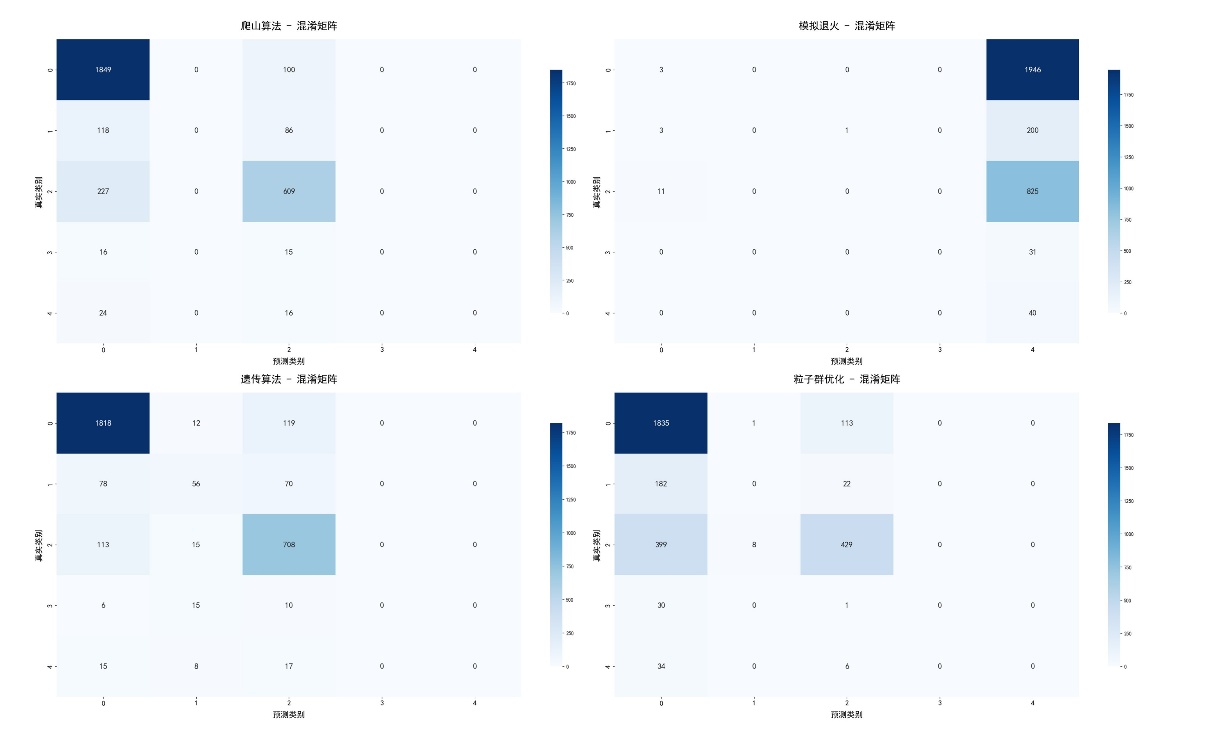


图 2 四种优化算法在贫血分类问题上的性能对比分析。遗传算法整体表现最优，在类别0（无贫血），对类别2（铁缺乏性贫血）识别率高，但对类别1（HGB贫血）、3（B12缺乏）和4（叶酸缺乏）的识别效果不佳。爬山算法表现次之，展现出与遗传算法相似的模式。粒子群优化算法表现模式相近但准确率较低。模拟退火算法整体表现最差，仅在类别0和类别3上显示出一定的识别能力。反映出类别不平衡问题对算法性能的显著影响。

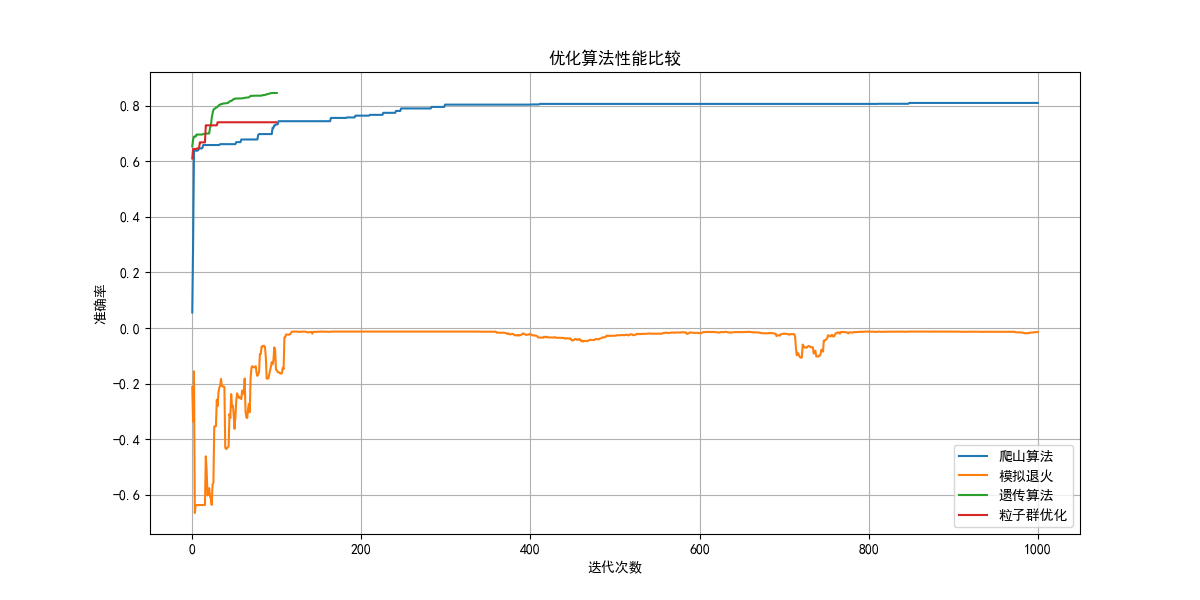


图 3 四种优化算法的收敛性能对比。遗传算法（绿线）表现出最快的收敛速度和最高的准确率，在100代后达到稳定，爬山算法（蓝线）展现出稳定的收敛趋势，最终达到约81%的准确率，粒子群优化（红线）收敛速度较快但性能略低，最终稳定在约71%；模拟退火算法（橙线）表现最不稳定，在训练初期出现剧烈波动，之后虽然趋于平稳但准确率维持在较低水平（约23%）。这一对比清晰地展示了遗传算法在该分类问题上的收敛速度，稳定性，和性能的优越性。

* 1. 实验说明

引用

[1] ALTENBERG L. The schema theorem and Price's theorem[M]. Foundations of genetic algorithms. Elsevier. 1995: 23-49.

[2] DEEP K, THAKUR M. A new crossover operator for real coded genetic algorithms[J]. Applied mathematics and computation, 2007, 188(1): 895-911.

[3] DEEP K, THAKUR M. A new mutation operator for real coded genetic algorithms[J]. Applied mathematics and Computation, 2007, 193(1): 211-30.

[4] GOLDBERG D E, DEB K. A comparative analysis of selection schemes used in genetic algorithms[M]. Foundations of genetic algorithms. Elsevier. 1991: 69-93.

[5] HE Y, HUI C-W. A binary coding genetic algorithm for multi-purpose process scheduling: A case study[J]. Chemical engineering science, 2010, 65(16): 4816-28.

[6] KATOCH S, CHAUHAN S S, KUMAR V. A review on genetic algorithm: past, present, and future[J]. Multimedia tools and applications, 2021, 80: 8091-126.

[7] KILIÇARSLAN S, CELIK, METE, ŞAHIN, SAFAK. Hybrid models based on genetic algorithm and deep learning algorithms for nutritional Anemia disease classification[J]. Biomedical Signal Processing and Control, 2022, 63.

[8] LI L, XUE-YUN C. Reconfiguration of distribution networks based on fuzzy genetic algorithms[J]. Proceedings-Chinese society of electrical engineering, 2000, 20(2): 66-9.

[9] LIN Y-C, HWANG K-S, WANG F-S. A mixed-coding scheme of evolutionary algorithms to solve mixed-integer nonlinear programming problems[J]. Computers & Mathematics with Applications, 2004, 47(8-9): 1295-307.

[10] PAVEZ-LAZO B, SOTO-CARTES J. A deterministic annular crossover genetic algorithm optimisation for the unit commitment problem[J]. Expert Systems with Applications, 2011, 38(6): 6523-9.

[11] UMBARKAR A J, SHETH P D. Crossover operators in genetic algorithms: a review[J]. ICTACT journal on soft computing, 2015, 6(1).

[12] YANG X, YANG Z, YIN X, et al. Chaos gray-coded genetic algorithm and its application for pollution source identifications in convection–diffusion equation[J]. Communications in Nonlinear Science and Numerical Simulation, 2008, 13(8): 1676-88.

[13] 白晓明, 韩家新. 差分遗传算法收敛性研究[J]. 软件导刊, 2016, 15(10): 26-9.

[14] 陈峰, 武小悦. 基于定向变异算子的求解GA欺骗问题研究[J]. 系统工程与电子技术, 2009, 31(01): 204-7.

[15] 陈辉, 张家树, 张超. 实数编码混沌量子遗传算法[J]. 控制与决策, 2005, 20(11): 1300-3.

[16] 戴朝华, 朱云芳, 陈维荣. 云自适应遗传算法[J]. 控制理论与应用, 2007, 24(4): 646-50.

[17] 金晶 苏. 一种改进的自适应遗传算法[J]. 计算机工程与应用, 2005, (18): 64-9.

[18] 梁旭, 王佳, 黄明. 解决大规模生产调度问题的一种新编码方法[J]. 计算机集成制造系统, 2008, 14(10): 1974-7.

[19] 彭茜, 王爱娟, 李峻阳, et al. 基于高效遗传算法的电网需求侧调度优化研究及其收敛性分析[J]. 电力系统保护与控制, 2022, 50(06): 33-42.

[20] 乔家庆, 付平, 孟升卫. 基于个体差异的遗传选择算子设计[J]. 电子学报, 2006, (S1): 2414-6.

[21] 郑朝晖, 张焱, 裘聿皇. 一种基于复数编码的遗传算法[J]. 控制理论与应用, 2003, 20(1): 97-100.

[22] HOLLAND J H. Genetic algorithms[J]. Scientific american, 1992, 267(1): 66-73.

[23] 马永杰, 云文霞. 遗传算法研究进展[J]. 2012.

[24] 张铃, 张钹. 遗传算法机理的研究 Ξ[J]. 2000.