

“复杂系统典型应用问题的建模与优化方法”结课报告

一、遗传算法在气体传感器阵列优化中应用的简要概述

遗传算法是一种基于种群的全局搜索优化技术。根据达尔文自然选择原理，遗传算法的目标是在由多个可行解组成的搜索空间中找到一个具有最佳适应度值的个体。这些解以基因型表示，适应度值越高的基因型，下一代被选择为亲本的概率越大。交叉和变异操作有助于算法有效地收敛，因此与其他方法相比，该算法获得局部最优解的机会更小。此外，遗传算法不需要任何领域的知识，也不需要任何有关优化问题结构的信息。即使目标函数不是光滑的，也能很好地应用遗传方法。因此，遗传算法可以作为一种有用的特征选择工具，和一种结构简单的、通用性强的算法，在气体传感器领域有着广泛的应用。这些应用中包含了对传感器的优化和对气体识别算法的优化：对传感器的优化着重体现在对传感器敏感材料结构或成分的优化，以及对传感器阵列组合的优化；对气体传感算法的优化，则是指使气体传感器能够更有效的分辨出待测气体的浓度，甚至对待测气体的成分进行分析。当然对于传感器阵列组合来说，对传感器阵列中传感器种类的优化，也是对气体分析能力的优化，也可以理解为是从另一方面优化了气体传感算法的参数。

为减小工作量，本文仅针对 2017 年至 2019 年最新的同时与气体传感器和遗传算法相关的文章进行阅读和信息提取。共收集了 5 篇相关文献，结合研究方向，以复现其算法为目的，精读了最新的一篇文献，略读了其余文献。这 5 篇文献中有两篇研究的是传感器阵列组合的优化问题，有两篇是针对气体传感算法的优化，有一篇是针对传感器几何结构的优化，所有问题均不同程度的将遗传算法应用其中。以下是这些，2017 年至 2019 年，遗传算法在气体传感器优化中应用实例的简要概述。

对于传感器阵列组合的优化。传感器阵列组合，是利用不同气体传感器间的传感特性差异，通过综合各传感器的响应数据，对未知成分气体的构成和各组分浓度进行预测的一种气体传感装置。这种传感装置应用的方面非常广泛，本次检索到的两篇相关文献的其中一篇，的应用背景是医学领域，内容为用于伤口感染细菌检测的气体传感器阵列优化^[1]，另一篇文献中也提到，在食品生产的质量控制中，这中传感器组合也被广泛应用^[2]。对于遗传算法在传感器阵列组合优化上

的应用，充分发挥了遗传算法不依赖于目标函数的导数/高阶导数进行搜索的特点。两篇文章的遗传算法设计主要区别在于染色体的设计。用于伤口细菌检测的气体传感器阵列优化，采用的是 0/1 编码，设置一个与传感器数据库相同长度的位串，每一位代表一个传感器，每一位的值对应着传感器是否被选上，0 代表未选，1 代表选择。这种选择方式的优点在于，传感器阵列的大小由遗传算法运算决定，有助于对传感器阵列中传感器数量的优化，以及在不限传感器数量大小的前提下寻找最优组合。另一篇文献对传感器阵列组合的优化中，则采用了确定的阵列大小，这意味着染色体对应的阵列长度是确定的，每个染色体上的基因的值代表着传感器库中的一个传感器，同时为了方便保证种群中每个染色体的唯一性，染色体中基因采用由小到大的排列方式，这种染色体适用于传感器大小受到限制的应用场景。在应用于医学领域的文献中，并没有详细叙述交叉变异的具体方法，同时该篇文献还指出相较于 Wilks lambda statistic (Wilks Λ -statistic)、Mahalanobis distance、linear discriminant analysis (LDA) 算法，遗传算法的计算时间消耗较大。另一篇文献对于算法的具体实现过程描述得像相当详细，因此本文尝试性的对其算法进行了复现，因此文献中遗传算法的具体执行过程，将在第二节中给出。

对于针对气体传感算法的优化。两篇文献分别测试了非支配排序遗传算法 (NSGA) 加卷积神经网络 (CNN) 的组合^[3]、以及遗传算法 (GA) 和模糊支持向量机 (FSVM) 的组合^[4]。并根据实际应用场景，设计了不同的染色体结构和交叉变异方法。前者的染色体设计较为常规，数据类型为列表，每一个列表元素代表着一个需要优化的参数，适值函数是重构误差和编码预测误差的和表示，优化目标为误差和最小，变异方法没有具体说明，只给出了以下参数：Mutation factor (M)、Mutation rate、Mutation probability、Crossover probability。后一篇文献则详细叙述了染色体的形式、交叉变异方法、适值函数形式。后者每个染色体拆分成了两个部分，分别对应了后续运算中所需的特征数量和对应参数分别称为：特征掩码、参数，各个部分都采用 0/1 编码，对于特征掩码部分，0/1 分别代表特征的缺失和存在，对于参数部分则有一个专门的解码函数。交叉变异方法对于这种染色体也需要特殊设计。对于交叉方法，特征掩码部分和各个参数部分需要分别进行单切点交叉。对于变异方法，特征掩码部分采用选择两个基因对调的方式进行变异，参数部分则采用改变被选择基因的 0/1 值的方式进行变异。对于适值函数，则将 FSVM 的准确性作为评价标准。

对于传感器几何结构的优化。这里选的文献中，遗传算法解决的问题是将光声光谱气体传感器中光声池的结构参数作为决策变量，将光声池品质因子和池常数作为目标函数值的多目标优化问题^[5]。和许多多目标优化问题一样，这里的光声池品质因子和池常数也是相互矛盾的量，一个增大另一个会减小。因此，运算结果是一个帕累托最优解前沿。对于具体计算过程，文献中仅提到采用 matlab 多目标遗传算法工具箱进行求解，采用二进制编码，以及给出的交叉率、变异率、种群大小和遗传代数，算法细节并没有做详细说明。不过，结论中提到，遗传算法的应用使得传感器性能得到了显著提升。

二、遗传算法在金属有机骨架阵列（MOF）甲烷传感器上的应用实例

本节是对文献 2 的复现，同时在复现过程中对原问题的适值函数进行了简化处理。具体原因是因为不能有效地得到原问题仿真结果，因此使用原问题的适值函数可能不能保证能得出有意义的求解结果。简化前的适值函数将在 2.1 小节中给出。

2.1. 遗传算法软件设计

具体软件设计，参考文件“..\GA_for_MOF_Array_opt\开发文档.docx”（Github 仓库目录）或“..\开发文档.pdf”（上交的作业 ZIP 文档）。

前文中提到在复现过程中对示值函数进行了简化，此处将对原适值函数进行简单说明。原始适应值函数的计算过程分为以下两大步，每大步包含若干小步：

第一大步：数据准备

● 第一步：生成待仿真样本点

具体过程：这里将待仿真气体的摩尔分数（mole fraction）分别称为 x 、 y 、 z ，三个参数分别代表氧气、氮气、甲烷的摩尔分数。这三个摩尔分数满足条件：三者和为 1。对每一个参数取 0 到 1 之间的值，步长为 0.2。

● 第二步：对 MOF 进行仿真

具体过程：对每种 MOF 使用上一步生成的待仿真样本点进行仿真，得到吸附质量。这会生成吸附值变量，变量类型为 list[list]，大小为 MOF 数量×样本点数量。

● 第三步：生成吸附值变量中每个样本点对应截断正态分布的概率密度曲线

具体过程：取吸附值变量中的每一个点，将其作为截断正态分布的均值，截断正态分布上下限分别设置为：0, $+\infty$ ，截断正态分布的标准偏差为均值的 10%。

- 第四步：生成每个 MOF 的摩尔分数-离散概率分布图

具体过程：首先，将第三步生成的所有截断正态分布概率密度曲线，取均值 $\pm 1\%$ 均值范围的概率值。第二步，对每个 MOF 对应的所有样本点取得的概率值归一化。得到的便是每个 MOF 对应的摩尔分数-离散概率分布图，数据大小为 MOF 数量 \times 样本点数量。

第二大步：面向染色体的适值计算

- 第一步：取染色体中包含的所有 MOF 做联合概率（joint probability）

具体过程：首先，去除染色体包含 MOF 的所有摩尔分数-离散概率分布图。然后，将这些摩尔分数-离散概率分布图按照对应采样点，逐点相乘。最后，将得到的值归一化，便得到了染色体对应 MOF 阵列的摩尔分数-离散概率分布图。

- 第二步：计算 MOF 阵列的对应三种气体的摩尔分数-离散概率分布图

具体过程：首先，将得到的 MOF 阵列的摩尔分数-离散概率分布图，分别以前面提到的 x 、 y 、 z 为横坐标，做成三张图。这样每张图的每个横坐标将对应多个点，将每个横坐标对应的点加和，最终得出三张横坐标纵坐标一一对应的图。这些图即 MOF 阵列的对应三种气体的摩尔分数-离散概率分布图，按照之前气体摩尔分数的取值，每张图有 51 个数据点。

- 第三步：计算 MOF 阵列对应三种气体的 KLD 值

具体过程：这里为方便表示 KLD 值计算方法，以 matlab 代码给出：

MATLAB code: $N = 51$; $Q_i = 1/N$; $KLD = \text{sum}(P \cdot \log_2(P/Q_i))$;

这里注意， N 为数据点个数， P 为每种气体的摩尔分数-离散概率分布图。最终得到 KLD_x 、 KLD_y 和 KLD_z 三个 KLD 值，分别对应三种气体。

- 第四步：求取当前染色体适值

具体过程：将 MOF 阵列算出的三个气体的 KLD 值乘积。

MATLAB code: $\text{fitness} = KLD_x * KLD_y * KLD_z$

2.2. 遗传算法的输出结果分析

通过对原问题代码的重构，可以看出该文献包含遗传算法的特点。最明显的特点有两条：

1、以选取父代的方法进行个体淘汰，每次循环产生的子代不参与本次循环的个体淘汰，直接进入下一循环；

2、产生变异的个体为交叉方法产生的每个子代，父代不参与变异操作。

在进行结果分析前，需进行以下两点说明：

1、输出文件说明：

- `Current_mof_SimResults.csv`: 当前采用的每个 MOF 对应适值;
- `Fitness_in_generations.csv`: 记录每次遗传算法执行的代数、每代的最大适值、每代的平均适值;
- `generations.csv`: 记录每次遗传算法每一代的所有个体;
- `mutation_rate.csv`: 此次执行遗传算法所用的变异率;
- `Theoretical maximum.csv`: 适应值的理论最大值。

2、遗传算法关键参数设置

表 1 遗传算法关键参数设置

参数	值	数据类型	参数	值	数据类型
种群大小	20	int	变异率	见“data”	int
染色体大小	5	int	MOF 列表	1,2,...,50	list
精英幸存者	20/5	int	MOF 对应值	读取文件	list
正常幸存者	20/5	int	迭代代数	50	int

表 1 中 MOF 对应值即上面提到的“Current_mof_SimResults.csv”文件中的值,需要读取到软件中。变异率,为 0.1 到 0.9 执行参数化扫描,步长 0.2。另外,还进行了一次针对交叉+变异同时进行的变异率为 0.01 到 0.1 的参数化扫描,步长 0.01。

数据输出结果分析:

从图 1 可以看出,随着变异率的增大,最终结果距离理论最大值越来越远,推测部分原因是变异率增大使得基因稳定性变差,优秀基因不容易保持。

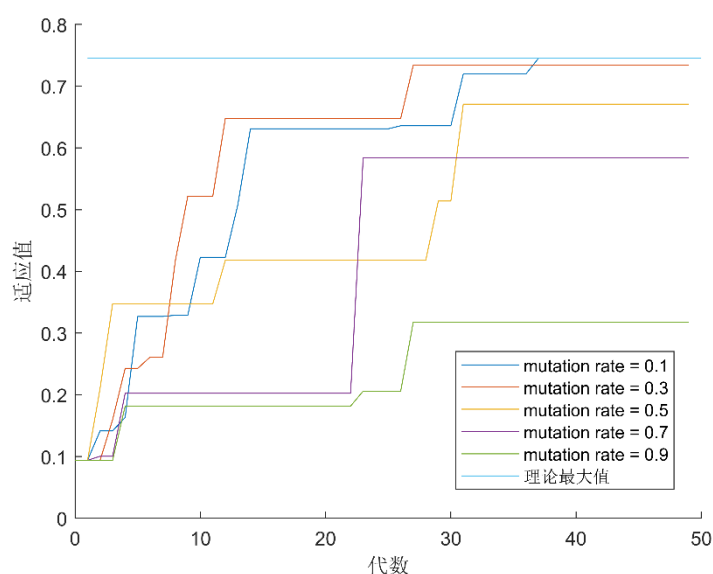


图 1 使用交叉+变异方法时在不同变异率下种群最大适值随迭代次数变化图

上述推论可以从图 1 中紫色线上得到一定程度的印证。具体的解释，是因为备选 MOF 数量为 50，阵列大小为 5，组合数实在太多，对于新产生的子代，因为变异率太大，不容易保持亲代的优良基因，进而导致进化缓慢。

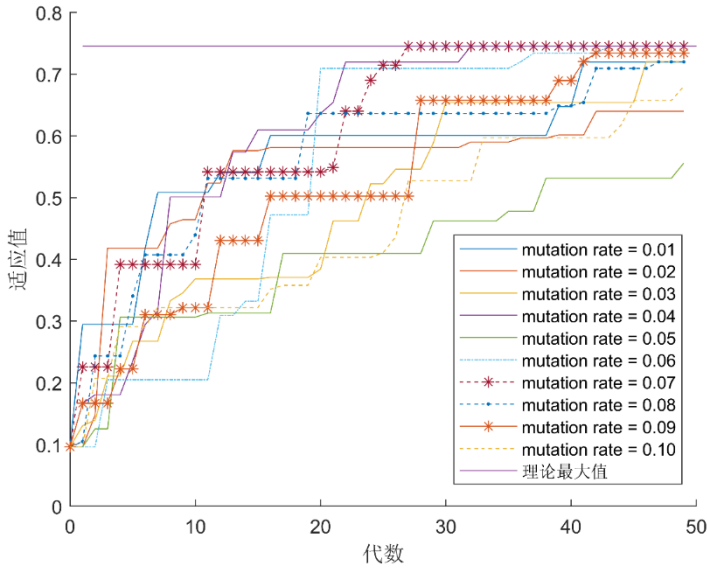


图 2 使用交叉+变异方法时在不同变异率下
种群最大适值随迭代次数变化图

如图 2 所示，当变异率较低的情况下，适应值增大的情况相对于图 1 要多。但是变异率过小也有一定风险。这是由这个遗传算法的淘汰机制导致的。因为，这里使用的交叉机制不会产生新的基因，淘汰机制会导致种群中的基因多样性越来越差，最终导致即便把种群中最好的基因凑一起也不会取到理论最大值。从图上看 0.01 到 0.1 的变异率实际差别不大。

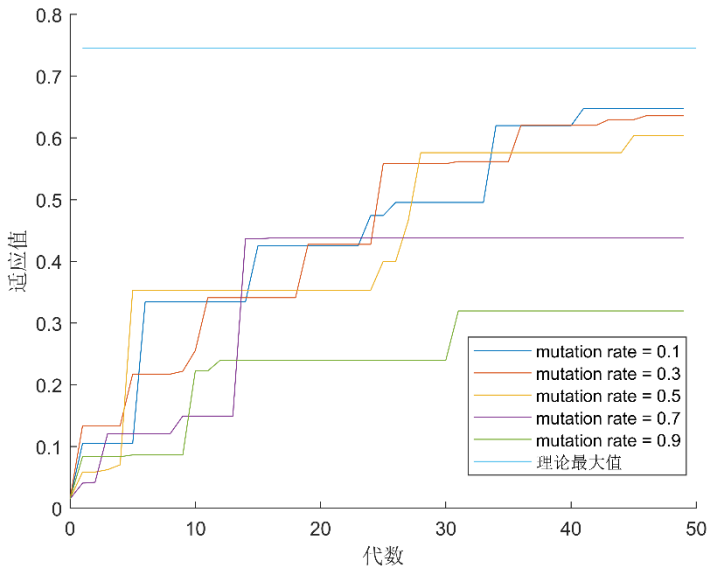


图 3 只使用变异方法时在不同变异率下种群最大适值随迭代次数变化图

图 3 为只使用变异方法产生子代时的情况，同样因为淘汰机制的原因，使得它的情况和图 1 相似。只是由于图 3 中少了交叉方法，使得算法不能把现有种群中的优良基因组合，进而达到理论最优值的效率慢一些。

三、总结

经过此次学习、查询资料和实践，让我更加充分的认识到了自己能力的不足，明确了在 `python` 学习中的方向，同时对于遗传算法有了更加充分的认识。对于不同的应用场景需要设计不同的遗传算法，遗传算法各环节的设计会因为具体问题的差异而千差万别，这充分体现了遗传算法对于应用场景有着强大的适应能力，也让我看到了智能优化算法在自己研究方向上的应用潜力。对于遗传算法的学习和编程语言的学习，实践永远是非常重要的，同时也要注意代码规范，尽量编写可以复用的代码，多做注释，不断练习不断提高，对语言和算法的理解也会越来越深。

四、参考文献

- [1] Sun H, Tian FC, Liang ZF, et. al. Sensor Array Optimization of Electronic Nose for Detection of Bacteria in Wound Infection[J]. IEEE Transactions on Industrial Electronics, 2017, 64(9): 7350-7358.
- [2] Gustafson J A, Wilmer C E. Intelligent Selection of Metal-Organic Framework Arrays for Methane Sensing via Genetic Algorithms[J]. ACS SENSORS, 2019, 4(6): 1586-1593.
- [3] Essiet I, Sun YX, Wang ZH. Big data analysis for gas sensor using convolutional neural network and ensemble of evolutionary algorithms[A]. Procedia Manufacturing[C]. Netherlands: ELSEVIER SCIENCE BV, 2019.629-634.
- [4] Liu TP, Zhang WT, McLean P, et. al. Electronic Nose-Based Odor Classification using Genetic Algorithms and Fuzzy Support Vector Machines[A]. IEEE International Conference on Systems, Man, and Cybernetics (SMC)[C]. USA: SPRINGER, 2018. 1309-1320.
- [5] Cheng G, Cao Y, Liu K, et. al. Numerical calculation and optimization of photoacoustic cell for photoacoustic spectrometer[J]. ACTA PHYSICA SINICA, 2019, 68(7): 074202.