上海师范大学

题目： 抗乳腺癌候选药物的优化建模

院（系、所） 信息与机电工程 专业 计算机科学与技术

课 程 科 目 最优化理论 第 二 学期

小组成员 廖山川、刘嘉俊、王一之

2023 年 6 月 3 日

摘　要

乳腺癌(breast cancer)是女性最常见的癌症类型，在不断增长的乳腺癌发病形势下，寻找特效药物对乳腺癌的治疗变得越来越重要。雌激素受体α亚型（Estrogen receptors alpha，ERα）被认为是治疗乳腺癌的重要靶标，在药物研发中，通常会建立化合物的活性预测模型来筛选潜在的活性化合物。本文根据给出的一系列化合物的分子描述符与活性数据，建立起化合物的 IC50 和 pIC50 值的预测关系，并建立起分子描述符与 ADMET 性质之间的关系。在此基础上，我们使用遗传算法对于化合物的 ADMET 性质进行优化，最终得到了模型最优值以及相对应的 20 个分子描述符。对其预测生物活性和 ADMET 性质，能够满足生物活性值 pIC50 值较高，同时 4 个 ADMET 值较好。

关键词：乳腺癌、药物预测、SVM、遗传算法

目 录

1. 问题背景与问题重述

1.1 问题背景

乳腺癌(breast cancer)是女性最常见的癌症类型。根据 2018 年发布的全球肿瘤统计报告，在全球女性恶性肿瘤占比中，乳腺癌新发病例及死亡率均列首位。全球约有 210 万新确诊的女性乳腺癌病例，占所有女性癌症病例近四分之一；死亡约 62.7 万例，女性全部恶性肿瘤死亡总数的 15.0%。因此，在现有临床药物的基础上，开发全新药物对乳腺癌的治疗具有深远的意义。

2019 年，我国国家癌症中心发布了最新的癌症统计数据，在我国十大高发恶性肿瘤中，乳腺癌位居女性肿瘤首位。数据统计显示，女性发病首位为乳腺癌，每年发病约为 30.4 万，占女性全部恶性肿瘤发病的 17.10%，较前一年的 16.51%呈增长态势，说明我国癌症的发病形势严峻，如图 1.1 所示。其他主要高发恶性肿瘤依次为肺癌、结直肠癌、甲状腺癌和胃癌等。

在不断增长的乳腺癌发病形势下，寻找特效药物对乳腺癌的治疗变得越来越重要。目前，临床上使用不同的分子靶向疗法来治疗不同类型的乳腺癌，在众多的治疗靶点中，雌激素受体(Estrogen receptors，ER)作为乳腺癌内分泌疗法的主要靶点，在超过 70%的乳腺癌患者中过度表达，并在乳腺癌的发展中扮演了主要的角色。ER 又分为 α 和 β 两种亚型，其中 ERα 主要位于乳房和子宫组织，而 ERβ 主要与神经系统有关，因此 ERα 被认为是乳腺癌的重要靶标。

综上，ERα 是乳腺癌的重要靶标，能够拮抗 ERα 活性的化合物可能就是治疗乳腺癌的候选药物，所以本文研究一系列化合物的生物活性值对靶标(ERα)的影响来寻找可能成为抗击乳腺癌发病的候选药物。

1.2 问题重述

该问题针对乳腺癌治疗靶标 ERα 提供了 1974 个化合物对 ERα 的生物活性数据。第一个文件“ERα_activity.xlsx”包含 3 列，第一列提供了 1974 个化合物的结构式，用一维线性表达式 SMILES（**S**implified **M**olecular **I**nput **L**ine **E**ntry

System）表示；第二列是化合物对 ERα 的生物活性值（用 IC_{50} 表示，为实验测定值，单位是 nM，值越小代表生物活性越大，对抑制 ERα 活性越有效）；第三列是将第二列 IC_{50} 值转化而得的 pIC_{50}。第二个文件“Molecular_Descriptor.xlsx”的 training 表（训练集）中，给出了上述 1974 个化合物的 729 个分子描述符信息（即自变量）。而第三个文件“ADMET.xlsx”的 training 表（训练集）中，提供了上述 1974 个化合物的 5 种 ADMET 性质的数据，其具体含义在表 1 中说明。

表 1 ADMET 性质

数据集	说明
Caco-2	‘1’代表该化合物的小肠上皮细胞渗透性较好，‘0’代表该化合物的小肠上皮细胞渗透性较差。
CYP3A4	‘1’代表该化合物能够被 CYP3A4 代谢，‘0’代表该化合物不能被 CYP3A4 代谢
hERG	‘1’代表该化合物具有心脏毒性，‘0’代表该化合物不具有心脏毒性
HOB	‘1’代表该化合物的口服生物利用度较好，‘0’代表该化合物的口服生物利用度较差
MN	‘1’代表该化合物具有遗传毒性，‘0’代表该化合物不具有遗传毒性

每个文件都包含训练集与相对应的测试集。

本文所要解决的优化问题是寻找并阐述化合物的哪些分子描述符，以及这些分子描述符在什么取值或者处于什么取值范围时，能够使化合物对抑制 ERα 具有更好的生物活性，同时具有更好的 ADMET 性质（给定的五个 ADMET 性质中，至少三个性质较好）。

以此为目标，我们需要先根据化合物的分子描述符与活性数据，建立起化合物的 IC50 和 pIC50 值的预测关系，并建立起分子描述符与 ADMET 性质之间的关系。根据建立起来的关系，通过优化算法对于化合物的 ADMET 性质进行优化，找出最佳的分子描述符。

2.模型建立

2.1 遗传算法模型介绍

根据具体优化问题，我们选择遗传算法来对 ADMET 性质进行优化。遗传算法（Genetic Algorithm, GA）是模拟达尔文生物进化论的自然选择和遗传学机理的生物进化过程的计算模型，是一种通过模拟自然进化过程搜索最优解的方法。

遗传算法有如下几个特点：

(1)直接对结构对象进行操作，不存在求导和函数连续性的限定

(2)具有内在隐并行性和更好的全局优能力

(3)采用概率化的寻优方法，不需要确定的规则就能自动获取和指导优化的搜索空间，自适应地调整搜索方向。

遗传算法以一种群体中的所有个体为对象，并利用随机化技术指导对一个被编码的参数空间进行高效搜索。其中，选择、交叉和变异构成了遗传算法的遗传操作；参数编码、初始群体的设定、适应度函数的设计、遗传操作设计、控制参数设定五个要素组成了遗传算法的核心内容。

一个基本遗传算法可以表示为：$SGA = (C, E, P_0, M, \phi, \Gamma, \psi, T)3$

其中：

C 表示个体的编码方案

E 表示个体适应度评价函数

P_0表示初始种群

M表示种群大小

ϕ表示选择算子

Γ表示交叉算子

ψ表示变异算子

T表示遗传算法终止条件

2.2 遗传算法具体步骤

1.染色体编码

(1)编码

解空间中的解在遗传算法中的表示形式。从问题的解(solution)到基因型的映射称为编码，即把一个问题的可行解从其解空间转换到遗传算法的搜索空间的转换方法。遗传算法在进行搜索之前先将解空间的解表示成遗传算法的基因型串(也就是染色体)结构数据，这些串结构数据的不同组合就构成了不同的点。

常见的编码方法有二进制编码、格雷码编码、浮点数编码、各参数级联编码、多参数交叉编码等。

(2)解码

遗传算法染色体向问题解的转换。

2.初始群体的生成

设置最大进化代数 T，群体大小 M，交叉概率 P_c，变异概率 P_m，随机生成 M 个个体作为初始化群体 P_0。

3.适应度值评估检测

适应度函数表明个体或解的优劣性。对于不同的问题，适应度函数的定义方式不同。根据具体问题，计算群体 P(t)中各个个体的适应度。

4.遗传算子

遗传算法使用以下三种遗传算子：

(1)选择

选择操作从旧群体中以一定概率选择优良个体组成新的种群，以繁殖得到下一代个体。个体被选中的概率跟适应度值有关，个体适应度值越高，被选中的概率越大。以轮盘赌法为例，若设种群数为 M，个体 i 的适应度为 f_i，则个体 i 被选取的概率为：

$$P_i = \frac{f_i}{\sum_{k=1}^{M} f_k}$$

当个体选择的概率给定后，产生[0,1]之间均匀随机数来决定哪个个体参加交

配。若个体的选择概率大，则有机会被多次选中，那么它的遗传基因就会在种群中扩大；若个体的选择概率小，则被淘汰的可能性会大。

(2)交叉

交叉操作是指从种群中随机选择两个个体，通过两个染色体的交换组合，把父串的优秀特征遗传给子串，从而产生新的优秀个体。

(3)变异

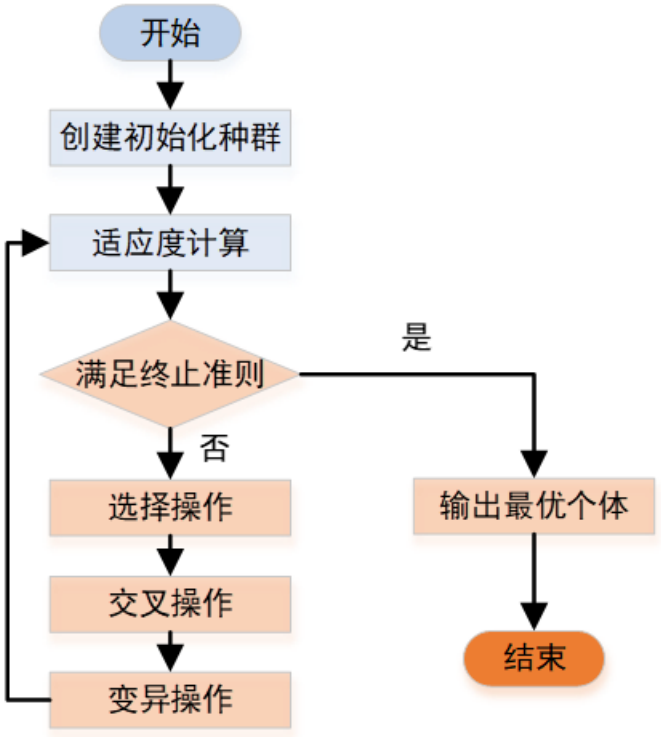
为了防止遗传算法在优化过程中陷入局部最优解，在搜索过程中，需要对个体进行变异，如单点变异，也叫位变异，即只需要对基因序列中某一个位进行变异，以二进制编码为例，即 0 变为 1 ，而 1 变为 0 。

5.终止判断条件

根据策略判断个体的适应度，是否符合优化准则。

最后：

M种群大小、 T终止条件、 P_c交叉概率、P_m*变异概率 这*四个运行参数需要预先设定。



3.实验与结果

3.1 步骤设计

（1）数据集预处理，按照设定的规则随机产生初始种群并设定遗传算法相关参数。

（2）将构建的初始种群分别带入化合物 pIC50 值预测模型和 ADMET 性质预测模型。

（3）开始遗传演化，依次计算种群中所有个体的适应度，优先选择适应度较高的个 体完成种群更新。

（4）判断迭代次数是否达到设定的最大迭代次数和优化目标，如达到则选出种群中最优个体，输出各变量值。否则，对种群继续进行遗传演化部分。

3.2 数据预处理

数据预处理流程图如图 4-1 所示。

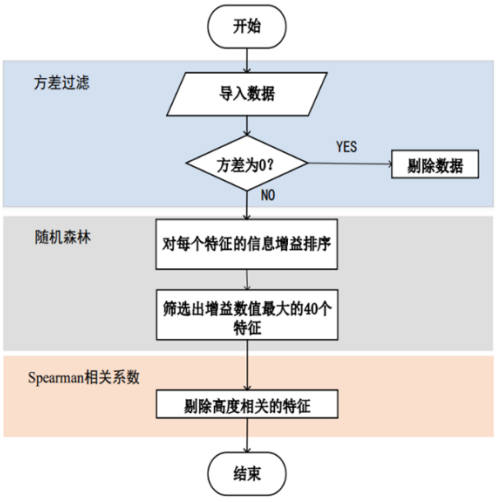


图 4-1 数据集预处理流程图

归一化：对每一列的分子描述符取值数据，进行离差标准化，映射到[0.1]区间；

方差过滤：计算表中 729 个变量的方差，筛选掉方差为 0 的变量；

随机森林：利用随机森林计算方差过滤后的特征变量对生物活性值的信息增益，对变量的贡献度进行排序，初步筛选出贡献度最高的前 40 个特征变量。

sprearman 相关系数：对选出的贡献度较高的 40 个变量做相关性检验，筛

除高相关性的变量,最终选出 20 个对生物活性具有显著影响的分子描述符。

sprearman 相关系数对两个变量之间的关联程度进行表示，sprearman 相关系数计算公式如下式所示：

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

20 个对生物活性具有显著影响的分子描述符如表 3-1 所示。

表 3-1 20 个对生物活性具有显著影响的分子描述符

变量名称	信息增益值	变量名称	信息增益值
MDEC-23	0.20686537246638967	nHBAcc	0.013027946401030333
LipoaffinityIndex	0.04362412409272269	nC	0.012522093437974615
C1SP2	0.04229182257822299	MLFER_A	0.011894134861576335
minsssN	0.034700586812533483	VC-5	0.010908139993956742
maxHsOH	0.03182560498124899	ATSc3	0.009872565844845372
maxssO	0.03114093818095146	SHsOH	0.009769723944909065
minHsOH	0.0275175713630276	TopoPSA	0.009003479973917535
minsOH	0.017187155444264676	minHBa	0.007409770955739792
BCUTc-1l	0.01695951737845333	MDEC-33	0.009003479973917535
minHBint5	0.016323367145658244	CrippenLogP	0.007295636815684211 5

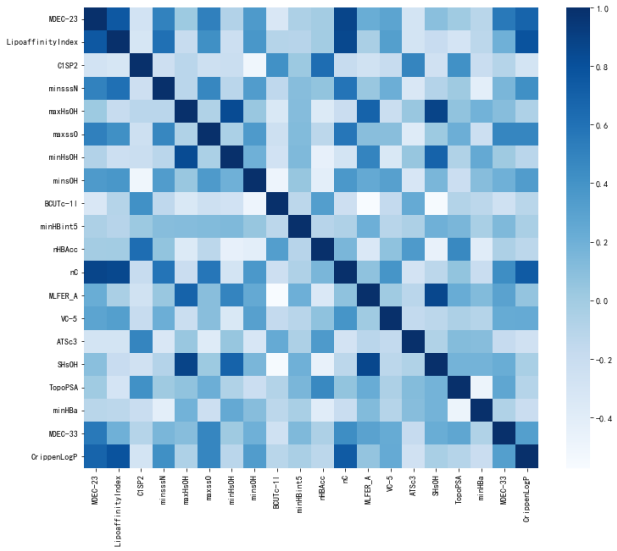


图 3-2 特征之间的相关系数热力图

由图 3-2 可以看出，我们选择的这 20 个信息增益值最大的特征变量之间相关关系颜色较浅，大部分变量线性相关关系不大，个别变量之间会存在一些较强的相关关系，属于可接受范围。

3.3 分子描述符取值范围

通过最后提交至 Kaggle 的分数可以看出 LightGBM 与 XGBoost 模型明显优于 Vanilla LSTM，从 LSTM 的角度来说，可能是由于 LSTM 获取的时间特征过少导致的，LSTM 实际参与训练的只有销量这一个特征，而并未能学习其余静态特征，这就导致模型并不能很好的拟合；从基于梯度提升树实现的集成算法——LightGBM 与 XGBoost 模型的角度来说，Boosting 集成多个决策树模型，每个模型都在尝试增强整体的效果，这就能使预测的准确率大大提高。

对数据预处理后确定的 20 个分子描述符，提取它们的最小值和最大值，如表 3-2 所示。

因此，设计的每个初始个体基因型均包含 20 个变量，其中各变量的取值范围要根据相应分子描述符的含义来确定，例如 C1SP2 代表双碳结合数量，nC 代表原子数，ndssC 代表原子性电拓形态，取值范围必须为整数。

表 3-2 取值范围表

变量名称	最小值	最大值	取值范围	变量名称	最小值	最大值	取值范围
MDEC-23	0	54.020151	R	nHBAcc	12.47	1207.82	R
LipoaffinityIndex	-4.58888	23.003681	R	nC	-0.372922	0.537745	R
C1SP2	0	6.7271374	R	MLFER_A	7	95	N
minsssN	0	2.7347721	R	VC-5	0	3.5326669	R
maxHsOH	0	20	N	ATSc3	1.46	5.75	R
maxssO	0	11.732336	R	SHsOH	-1.487697	113.56736	R
minHsOH	0	66	R	TopoPSA	0	28	N

minsOH	-1.198073	11.555324	R	minHBa	-1.059135	177.78781	R
BCUTc-1l	-0.856	7.754	R	MDEC-33	2.454098	13.665048	R
minHBint5	0	1.4810587	R	CrippenLogP	-3.592	14.283	R

3.4 适应度函数

综合 pIC50 值预测和 ADMET 性质预测，我们需要使得化合物的生物活性高的同时五个指标至少三个性质较好。五个指标的良好标准为[1,0,0,1,0]。因此将两者按照 1:1 的比例结合为目标函数$fitness$，其中 pIC 代表各化合物对 pIC50 的预测值，Co代表各化合物与[1,0,0,1,0]的匹配程度，例如[0,1,0,1,0]与标准的匹配程度为 3。$fitness$越低，该样本的适应度越高。

$$fitness = -（0.5*\frac{pIC}{10} + 0.5*\frac{Co}{5}）$$

根据上述公式我们首先要确定 pIC 和 Co 的值，我们的做法是通过训练这个 20 个分子描述符对 Erα 活性影响和 ADMET 性质的影响构建模型，从而通过遗传算法不断进行迭代。从而求得当这 20 个分子描述符取什么值的时候。化合物对抑制 ERα 具有更好的生物活性，同时具有更好的 ADMET 性质。

3.5 预测 ADMET 性质

本次训练预测 ADMET 性质的模型我们采用 SVM，SVM 在训练多分类任务的时候可以转换多个单标签分类问题，极大简化了模型的构建和训练，同时二进制编码标签可以增强模型的准确性和鲁棒性。单个 SVM 可能对多目标的预测效果较差，因此我们采用集成学习 SVM 去预测 ADMET 性质，通过参数优化最终达到理想的效果。

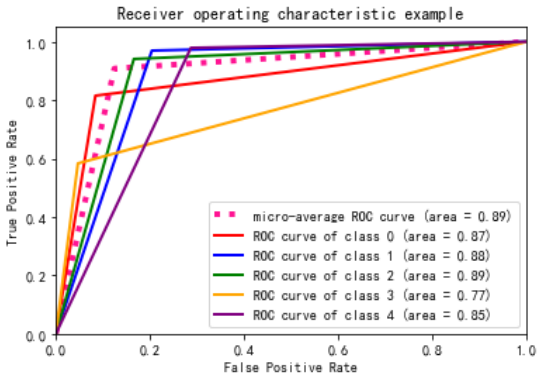


图 3-3 ROC 曲线和 AUC

由图 4-3 所示，选取最好结果的预测 ADMET 性质模型，对文件“ADMET.xlsx”的 test 表中的 50 个化合物进行相应的预测，得到部分结果如表 3-3 所示。

表 3-3 ADMET 性质预测表（前 10）

化合物	Caco-2	CYP3A4	hERG	HOB	MN
COc1cc(OC)cc(\C=C\c2ccc(OS(=O)(=O)[C@@H]3C[C@@H]4O[C@H]3C(=C4c5ccc(O)cc5)c6ccc(O)cc6)cc2)c1	0	1	1	0	1
OC(=O)\C=C\c1ccc(cc1)C2=C(CCOc3ccccc23)c4ccc(O)cc4	0	1	1	0	1
COc1ccc2C(=C(CCOc2c1)c3ccc(O)cc3)c4ccc(\C=C\C(=O)O)cc4	0	1	1	0	1
OC(=O)\C=C\c1ccc(cc1)C2=C(CCOc3cc(F)ccc23)c4ccc(O)cc4	0	1	1	0	1
OC(=O)\C=C\c1ccc(cc1)C2=C(CCSc3cc(F)ccc23)c4ccc(O)cc4	0	1	1	0	1
CC(=O)\C=C\c1ccc(cc1)C2=C(CCOc3cc(F)ccc23)c4ccc(O)cc4	0	1	0	0	1
Oc1ccc(cc1)C2=C(c3ccc(\C=C\c4ccccc4)cc3)c5ccc(F)cc5OCC2	0	1	0	0	0
Oc1ccc(cc1)C2=C(c3ccc(\C=C\C(=O)c4ccccc4)cc3)c5ccc(F)cc5OCC2	0	1	1	0	1
OC(=O)\C=C\C=C\c1ccc(cc1)C2=C(CCOc3cc(F)ccc23)c4ccc(O)cc4	0	1	1	0	1
CCN(CC)C(=O)\C=C\c1ccc(cc1)C2=C(CCOc3cc(F)ccc23)c4ccc(O)cc4	0	1	1	0	1

经相关对比，我们的预测 ADMET 性质模型准确率较高。

3.6 预测 pIC50 值

本次训练模型预测 pIC50 值采用回归树作为弱分类器，Adaboost 作为集成学习算法。

集成学习算法可以提高模型的准确度，鲁棒性，可扩展性。

我们采用网格搜索出最优参数。所得结果中，Mean Squared Error 为 0.5733006346609737，R-squared 为 0.7170331546960137。其中 R-squared 如图 3-

4 所示：

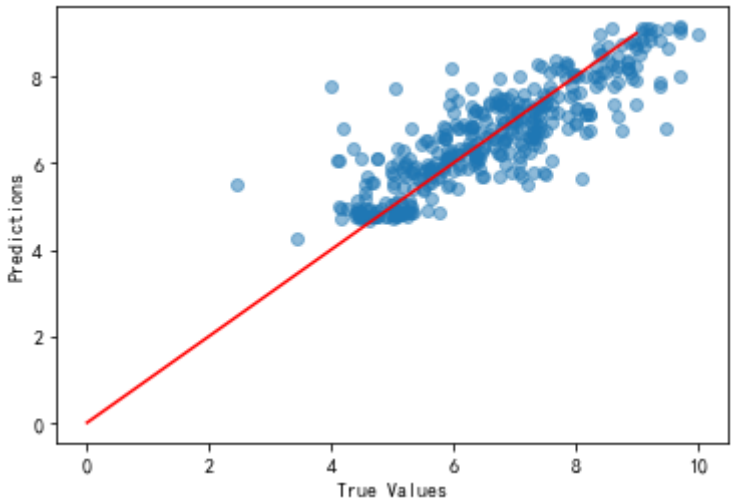


图 3-4 R2

将此预测 pIC50 值模型，对文件“ERα_activity.xlsx”的 test 表中的 50 个化合物进行相应的预测，得到部分结果如表 3-4 所示。

表 3-4 pIC_{50} 值预测表（前 10）

化合物	pIC_{50}
COc1cc(OC)cc(\C=C\c2ccc(OS(=O)(=O)[C@@H]3C[C@@H]4O[C@H]3C(=C4c5ccc(O)cc5)c6ccc(O)cc6)cc2)c1	8.089
OC(=O)\C=C\c1ccc(cc1)C2=C(CCOc3ccccc23)c4ccc(O)cc4	6.987
COc1ccc2C(=C(CCOc2c1)c3ccc(O)cc3)c4ccc(\C=C\C(=O)O)cc4	7.824
OC(=O)\C=C\c1ccc(cc1)C2=C(CCOc3cc(F)ccc23)c4ccc(O)cc4	7.251
OC(=O)\C=C\c1ccc(cc1)C2=C(CCSc3cc(F)ccc23)c4ccc(O)cc4	8.539
CC(=O)\C=C\c1ccc(cc1)C2=C(CCOc3cc(F)ccc23)c4ccc(O)cc4	7.206
Oc1ccc(cc1)C2=C(c3ccc(\C=C\c4ccccc4)cc3)c5ccc(F)cc5OCC2	5.882
Oc1ccc(cc1)C2=C(c3ccc(\C=C\C(=O)c4ccccc4)cc3)c5ccc(F)cc5OCC2	6.009
OC(=O)\C=C\C=C\c1ccc(cc1)C2=C(CCOc3cc(F)ccc23)c4ccc(O)cc4	6.970
CCN(CC)C(=O)\C=C\c1ccc(cc1)C2=C(CCOc3cc(F)ccc23)c4ccc(O)cc4	7.910

经相关对比，我们的预测 pIC50 值模型能较好地拟合实际值。

3.7 初始个体及种群

初始种群数量依次选择 10，20，3 … 100。最大迭代次数 10 次，其对应的适应度曲线如图 3-5 所示。

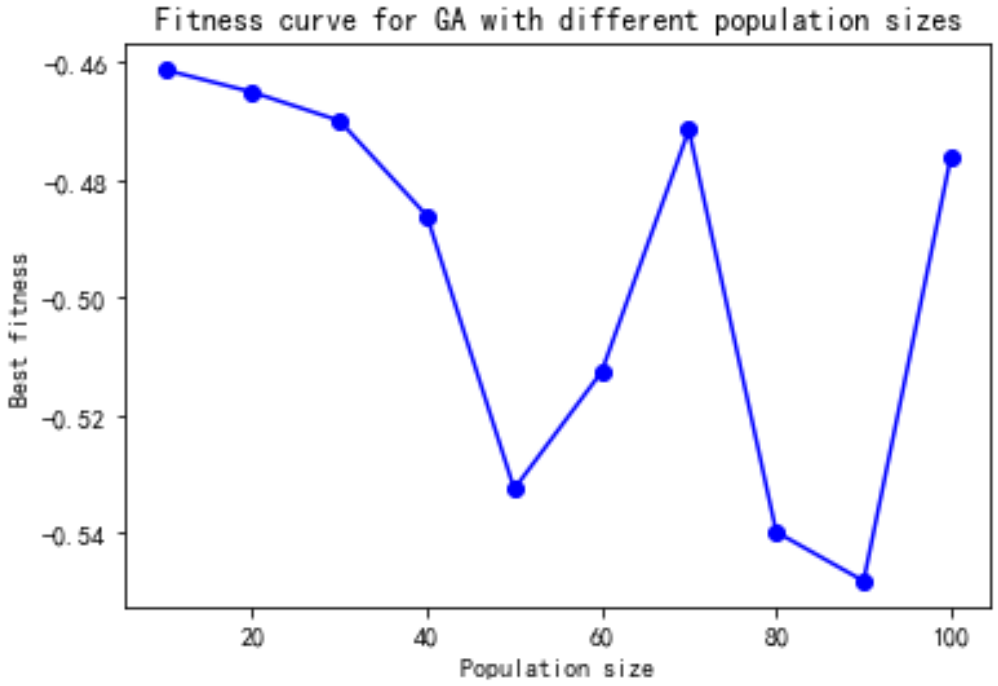


图 3-5 初始种群数量对应的适应度曲线（10 次迭代）

依据曲线变化程度及最后适应度情况，可以看出种群在 90 的时候达到最佳效果，因此最终确定种群数量为 90。

3.8 最大迭代次数

最大迭代次数依次选择 10，20，3 … 100。由以上确定的初始化种群为 90，其对应的适应度曲线如图 3-6 所示。依据曲线变化程度及最后适应度情况，最大迭代次数在 80 的时候达到最有效果，因此最终确定最大迭代次数为 80。

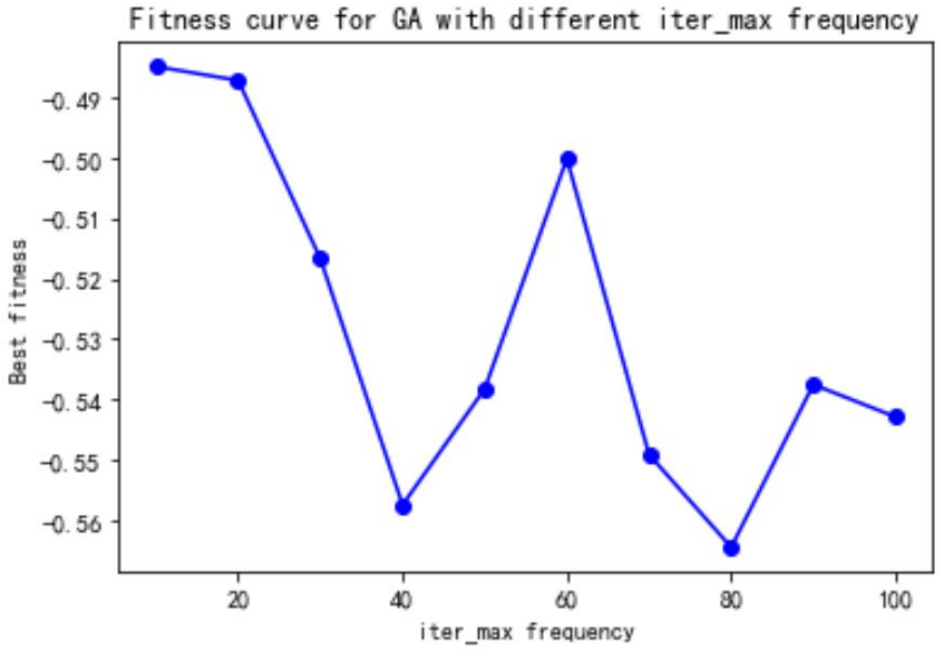


图 3-6 最大迭代次数对应的最后一代适应度曲线（初始种群 90）

依据曲线变化程度及最后适应度情况，最大迭代次数在 80 的时候达到最有效果，因此最终确定最大迭代次数为 80。

3.9 变异方法

我们对个体使用高斯突变进行变异操作，在进行变异时用一个均值 μ、方差为 σ2 的正态分布的一个随机数来替换原有基因值。高斯变异算子的公式如下所示。

$$\mathcal{X}_{i,j}^{\sigma} = \mathcal{X}_{1,j}^{\rho} + \sigma_{i,j} \cdot \mathcal{N}(0,1)$$

3.10 模型评估标准

$$fitness = -（0.5 * \frac{pIC}{10} + 0.5 * \frac{Co}{5}）$$

对上述公式，观察到 pIC50 的取值范围绝大多数在 0~10 范围内，因此可以得到前部分公式的变化范围为 0~0.5。当确保 ADMET 性质至少有三个符合标准时，后半部分公式的变化范围为 0.3~0.5。因此$fitness$的值小于等于-0.8 时，至少有三个 ADMET 性质得到保证，同时化合物的 ERα 生物活性也较高。于是

$$fitness \leq -0.8$$

是衡量最终模型是否可行的标准。

对于我们最终的遗传算法模型，迭代结束后得到最重要 20 个特征的最优解

如表 3-5 所示。

表 3-5 20 个特征的值

分子描述符	最终值	分子描述符	最终值
MDEC-23	0.80946415	nHBAcc	0.78604602
LipoaffinityIndex	0.67885272	nC	0.97718745
C1SP2	0.19770474	MLFER_A	0.56740144
minsssN	0.1234054	VC-5	0.13613332
maxHsOH	0.54991111	ATSc3	0.38090762
maxssO	0.96898621	SHsOH	0.93033319
minHsOH	0.03077406	TopoPSA	0.16428174
minsOH	0.85566383	minHBa	0.00686389
BCUTc-1l	0.09758825	MDEC-33	0.82642149
minHBint5	0.77185635	CrippenLogP	0.09316338

经过 pIC50 预测模型结果为 6.3594，即 pIC=6.3594，经过 ADMET 性质预测模型结果为 [1,0,0,1,0]，对照 ADMET 性质标准[1,0,0,1,0],可得性质优良数目为 5，即Co=5。最优适应度可计算为

$$fitness = -(0.5 * 0.63594 + 0.5 * 1.0) = -0.81797$$

因此我们的遗传算法模型达到了$fitness$的值小于等于-0.8 的标准，我们能找到合适的最重要的 20 个分子描述符的取值来保证化合物的 ERα 生物活性较高，同时满足较好的 ADMET 性质。

4.总结

4.1 优点总结

我们对建模问题的处理从两个方面建立预先的模型入手，再使用遗传算法综合寻找最优解，任务明确，思路较为清晰。

预处理阶段：首先对分子描述符的数据进行预处理，然后使用随机森林训练分子描述符对 Erα 的抑制效果得到 20 个影响因子最大的分子描述符。

获取 ADMET 预测模型：采用集成 SVM 算法训练 ADMET 预测模型，可以实现基于 20 个主要分子描述符来预测 5 个 ADMET 性质指标。该模型避免了使用单个 SVM 可能带来的错误分类现象。

获取 pIC50 值的预测模型：采用集成 Adaboost 算法训练化合物 IC50 值和对应的 pIC50 值的预测模型，可以实现基于 20 个主要分子描述符来预测 Erα 的抑制效果。而且 Adaboost 模型运行效率更高也更加灵活，也不会因为样本数据中某极个别异常数据从而影响到整个模型，稳定性较强。

最终求解：基于遗传算法不断迭代寻找出 20 个分子描述符的最佳取值。基于该问题提出的模型评估标准高度适配了优化模型，经过一定的迭代次数后能够准确、高效的得到 20 个变量的取值。既保证了较高的 ERα 生物活性，也使得 ADMET 性质得到了优化。

4.2 不足总结

对于预处理阶段，对特征筛选过程中根据信息增益值和变量间关系进行计算，属于定量分析。但实际药物生产过程中还会存在其他一些分子之间反应等问题，后续研究应将定量与定性相结合以对关键特征筛选。

对于获取 ADMET 预测模型阶段和获取 pIC50 值的预测模型阶段，集成 SVM 算法存在难以应对大规模训练样本的问题。同时，集成 Adaboost 算法对较小规模的特征变量拟合度较好，比如本课题的 20 个，但是若实际使用几百个，则性能严重不足。

对于最终求解，遗传算法的使用只能够确定相关变量的取值，而不是取值范围。同时该模型的迭代次数可能较少，无法得到更好的准确度。