

# 基于特征进化选择随机森林的 MCI 自动诊断

高峰 郑丽丽 顾进广

(武汉科技大学计算机科学与技术学院 湖北 武汉 430065)

(武汉科技大学大数据科学与工程研究院 湖北 武汉 430065)

(湖北省智能信息处理与实时工业系统重点实验室 湖北 武汉 430065)

(富媒体数字出版内容组织与知识服务重点实验室 湖北 武汉 430065)

**摘要** 近年来,作为正常与阿尔茨海默病过渡阶段的轻度认知障碍(Mild Cognitive Impairment, MCI)病症的研究备受关注。但目前的医学 MCI 人工诊断不仅参考的特征局限性较大,且依靠人工判定,易产生主观上的误差。因此,提出一种基于随机森林的 MCI 自动诊断方法,通过机器学习的方式,高效准确地判定 MCI;同时应用遗传算法更高效地搜索求解模型的最优参数。结果表明,该方法与医学人工诊断方式相比准确率提高约 5%,且在求取随机森林的最优参数问题上,与网格搜索相比,遗传算法所用时间约为其 1/45。

**关键词** MCI 随机森林 遗传算法 最优参数

**中图分类号** TP3 **文献标志码** A **DOI**:10.3969/j.issn.1000-386x.2024.06.037

## MCI AUTOMATIC DIAGNOSIS BASED ON FEATURE EVOLUTION AND RANDOM FOREST SELECTION

Gao Feng Zheng Lili Gu Jinguang

(College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, Hubei, China)

(Big Data Science and Engineering Research Institute, Wuhan University of Science and Technology, Wuhan 430065, Hubei, China)

(Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial, Wuhan 430065, Hubei, China)

(Key Laboratory of Content Organization and Knowledge Service for Rich Media Digital Publishing, Wuhan 430065, Hubei, China)

**Abstract** In recent years, research on the condition of mind cognitive impairment (MCI), which is the normal and excessive stage of Alzheimer's disease, has attracted much attention. However, the current medical MCI manual diagnosis not only has relatively large limitations in the referenced features, but also relies on manual judgment, which is prone to subjective errors. Therefore, this paper proposes an automatic diagnosis method of MCI based on random forest, hoping to determine MCI efficiently and accurately through machine learning. At the same time, in order to obtain the optimal parameters of the random forest MCI diagnosis model more efficiently, genetic algorithm was combined. The results show that the accuracy of this method is about 5% higher than that of medical manual diagnosis, and the time taken by genetic algorithm is shortened by nearly 45 times compared with grid search on the problem of obtaining the optimal parameters of random forest.

**Keywords** MCI Random forest Genetic algorithm Optimal parameters

## 0 引言

阿尔茨海默病(Alzheimer's disease, AD)是老年期

常见的慢性精神衰退性疾病,会对老年人的生活产生巨大影响,给个人、家庭和社会带来极大压力和痛苦。且 AD 病症发病率较高,到 2050 年时预估全球每 85 人就有一人罹患此病<sup>[1]</sup>,给家庭和社会带来了巨大的

压力。另外,现有针对AD病症的药物非常有限。但幸运的是,研究表明早期发现、预防治疗可以有效减缓AD的发病情况。轻度认知障碍(MCI)是介于正常老化和AD之间的过渡阶段<sup>[2]</sup>,MCI的患者是AD的高危人群,研究表明约44%的MCI患者在3年后转化为AD,平均年转化率为15%<sup>[3]</sup>,而每年仅有1%~2%正常老化转化为AD。因此,对于MCI患者的提前诊断和早期治疗具有重要价值。

本文提出一种基于随机森林(Random Forest, RF)算法的MCI自动诊断方法,并且通过遗传算法(Genetic Algorithm, GA)获取随机森林的最优参数,期望获得更好的MCI诊断效果。

## 1 相关工作

医学研究表明,MCI病症可以通过患者的基本情况、文化教育年数和简易智力状态检查(Mini-mental State Examination, MMSE)量表的判定分数等特征进行有效的判定<sup>[4]</sup>。依据文献<sup>[5]</sup>,医学MCI诊断方式如图1所示。

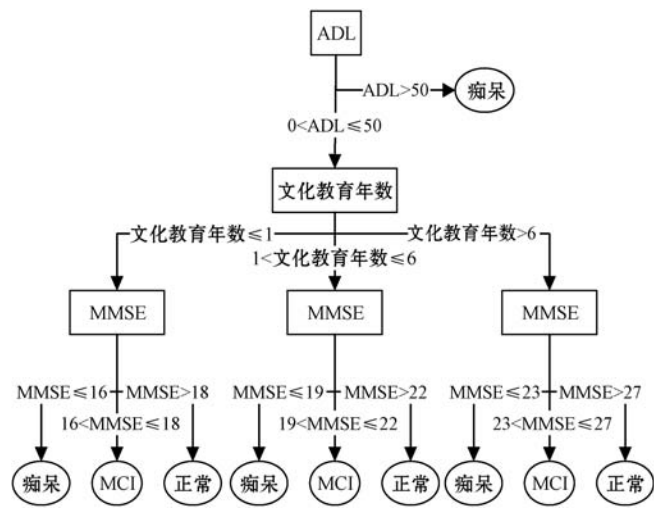


图1 医学MCI诊断方式

由图1可见,医学MCI判定的方式,对于MMSE、文化教育年数和ADL等特征过于依赖,而易忽视其他可能影响的特征。遇到情况较复杂的情况,利用以往的判定方法,可能会忽视重要因素,产生错误。且目前MCI主要是依靠医生判断,因该判断方式是人为过程,可能产生主观的误差。因此,针对以上问题,本文提出使用随机森林对MCI病症进行自动诊断,希望通过机器学习的方式高效准确地进行MCI的诊断。

同时,以往的随机森林求取最优参数时,通常采用网格搜索(Grid Search)方式<sup>[6]</sup>。这是一种通过穷举所有参数组合,逐一进行测试获取最优值的暴力求解方法,在时间和空间上都造成了很大的浪费。

本文采用遗传算法求取随机森林MCI诊断模型的最优参数,通过MCI样本特征设定随机森林参数范围后,根据参数范围设定染色体编码规则,生成参数种群,经过种群的世代演化,最终收敛于最优参数解。相比于网格搜索,遗传算法是一种基于生物自然选择与遗传机理的随机搜索与优化方法<sup>[7]</sup>,具有快速随机的全局搜索能力。

实验结果表明,本文方法不仅节省了宝贵医疗成本,减轻医生工作负荷,同时与目前的医学MCI诊断方式相比,准确率也有了显著提高。并且采用遗传算法获取随机森林最优参数过程的时间远低于网格搜索。

## 2 方法

### 2.1 随机森林

随机森林算法是一种基于决策树的集成学习(Ensemble learning)算法<sup>[8]</sup>,运用Bagging<sup>[9]</sup>思想并行构造多棵“好而不同”<sup>[10]</sup>的决策树分类器。随机森林主要用于处理分类和回归问题,对于处理类似MCI这种高维数据具有较好的性能<sup>[11]</sup>,图2为随机森林MCI诊断示例。

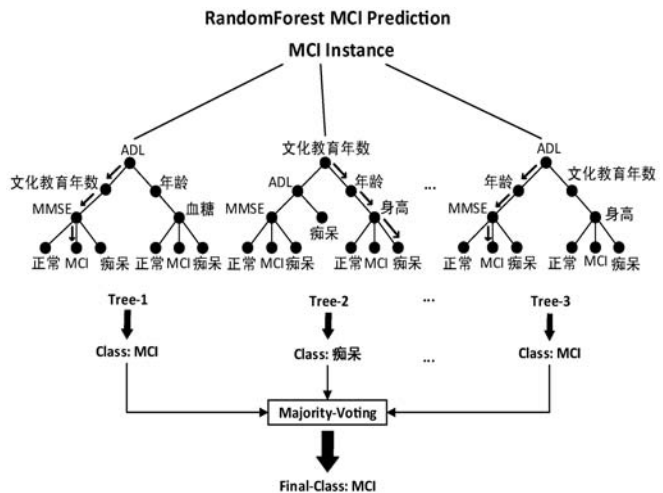


图2 随机森林MCI诊断示例

如图2所示,不同的决策树对于MCI的决策过程是不同的,例如Tree-1将ADL作为根节点决策特征,而Tree-2将文化年作为根节点特征,因此对于相同的MCI样本,不同决策树可能得出不同的预判结果。而随机森林对MCI的预判结果是由以上多棵决策树投票决定的,输出结果采用少数服从多数的原则,优化了单棵决策树的过拟合和复杂剪枝问题<sup>[12]</sup>。

随机森林的主要参数有: $n_{estimators}$ 、 $m_{ax\_depth}$ 、 $m_{ax\_features}$ 、 $m_{in\_samples\_leaf}$ <sup>[13]</sup>。其参数说明如表1所示。

表 1 随机森林 MCI 预判模型参数说明

参数名	含义	说明
$n_{estimators}$	决策树数目	一般决策树越多,模型精度越高,相对应的模型评估时间也越长
$m_{ax\_depth}$	决策树最大深度	决策树最大深度应该根据数据的特征值个数设定,本文 $m_{ax\_depth}$ 的设置应该处于 1 至 MCI 的特征数之间;其次,决策树最大深度设置过小达不到理想判定效果,但设置过大,会产生过拟合现象
$m_{ax\_features}$	最大特征数	挑选特征值作为节点特征时, $m_{ax\_features}$ 表示备选特征数,也不应该大于数据的特征数,本文实验中设定在 1 至 MCI 特征数之间
$m_{in\_samples\_leaf}$	叶子最小特征数	叶子节点对应的样本数小于该值时,就会与其直接父节点进行合并

## 2.2 遗传算法

遗传算法是模拟生物学中基因遗传进化规律的一种算法思想<sup>[14]</sup>,依据自然选择定律中的优胜劣汰原则,经过种群进化的模拟,快速搜索最优解。图 3 为遗传算法求取随机森林 MCI 诊断模型最优参数的流程。

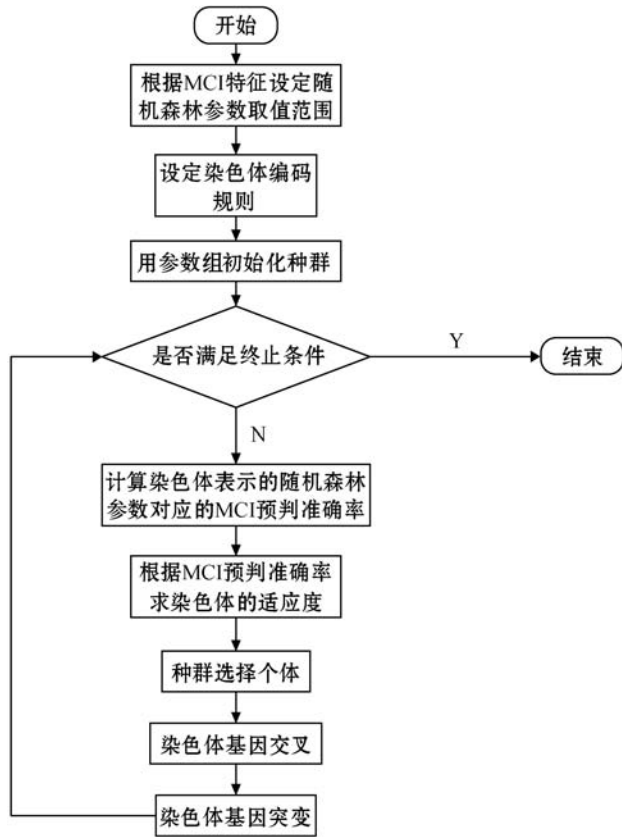


图 3 遗传算法求取随机森林 MCI 诊断模型最优参数流程

(1) 根据 MCI 特征设定随机森林参数范围。因随机森林的决策树个数一般在 100 后趋于稳定<sup>[15]</sup>,所以设定  $n_{estimators}$  的取值范围为  $[1, 121]$ ; 本文 MCI 的特征数为 22 个,因此设定  $m_{ax\_features}$  的取值范围为  $[1, 22]$ ; 为防止决策树出现过拟合现象,  $m_{ax\_depth}$  的取值应略小于 MCI 的特征数,故设定为  $[1, 16]$ ;  $m_{in\_samples\_leaf}$  的取值范围设定为  $[1, 8]$ 。

(2) 设定染色体编码规则。本文将随机森林 MCI

诊断模型的一组参数组合作为一个染色体,染色体也称为个体,一个染色体含有多个基因,一个基因用一个 2 进制数表示,一个参数由一至多个基因表示,基因的个数由参数的范围以及二进制编码与参数实值的对应规则而定,基因越多,精度越大。

根据以上得出的参数取值范围,设定基因个数为 14 个,  $n_{estimators}$ 、 $m_{ax\_depth}$ 、 $m_{ax\_features}$ 、 $m_{in\_samples\_leaf}$  可分别对应 4、4、3、3 个基因表示。例如,二进制编码和参数实值的对应规则如式(1)所示。

$$\begin{cases} n_{estimators} = 1 + B_{n_{estimators}} \times 8 \\ m_{ax\_depth} = 1 + B_{m_{ax\_depth}} \\ m_{ax\_features} = 1 + B_{m_{ax\_features}} \times 3 \\ m_{in\_sample\_leaf} = 1 + B_{m_{in\_samples\_leaf}} \end{cases} \quad (1)$$

根据以上 2 进制编码规则,例如,参数组合  $c_{chrom} = \{n_{estimators} = 89, m_{ax\_depth} = 7, m_{ax\_features} = 19, m_{in\_samples\_leaf} = 7\}$  对应的染色体可表示为 1011|0110|110|110。

(3) 用参数组初始化种群。利用上述方式生成一系列染色体表示随机森林 MCI 诊断模型参数组,形成一个种群,进行接下来的选择、交叉和变异过程<sup>[16]</sup>。种群中染色体数目记为  $c_{chrom\_num}$ 。初始化种群通常采用随机生成的方式。

(4) 通过 MCI 预判准确率求染色体的适应度。染色体的适应度对应了当前个体的优劣程度,一般由目标函数决定。本文个体适应度是根据个体对应的参数组在随机森林 MCI 诊断模型的准确率计算的,个体  $i$  的 MCI 诊断准确率对应为  $M_{Cl_{acc}(i)}$ ,种群中 MCI 预测准确率最小值表现为  $M_{Cl_{acc\_min}}$ 。首先计算本轮循环,种群中个体的 MCI 诊断准确率与 MCI 诊断表现最差的个体的准确率之差,记作个体  $i$  在种群中的表现差值  $f(i)$ ,  $f(i)$  计算如式(2)所示。

$$f(i) = M_{Cl_{acc}(i)} - M_{Cl_{acc\_min}} \quad (2)$$

接着计算该种群的表现差值之和,如式(3)所示。

$$F = \sum_1^{c_{chrom\_num}} f(i) \quad (3)$$

最后个体  $i$  的适应度  $fit(i)$  计算式如式(4)所示。

$$fit(i) = f(i) / F \quad (4)$$

(5) 种群选择个体。根据自然界优胜劣汰的选择规律,根据染色体的适应度对个体进行选择,即对随机森林 MCI 诊断模型的参数组合进行筛选。适应度大的个体更可能被保留下来,但是并不意味着适应度小的个体一定消失,其中存在一些个体虽然适应度不高,但可能包含优秀基因,作为之后的遗传发展基础。

本文采用轮盘赌选择法<sup>[17]</sup>,对个体进行选择。计算适应度累计值  $c_{umsum_i}$  作为个体  $i$  的选择条件,计算式如式(5)所示。

$$c_{umsum_i} = \sum_1^i fit(x) \quad (5)$$

根据上述所示,  $c_{umsum_i} \in [0, 1]$ , 因此在  $[c_{umsum_{min}}, 1]$  范围内,随机生成  $c_{hrom\_num}$  个选择算子  $m_s$ , 对  $m_s$  列表进行从小到大排序后,循环  $m_s$  列表,当  $m_{s_i} < c_{umsum_k}$  时保留个体  $k$ , 否则,  $m_{s_i}$  继续与下一个个体比较,轮盘赌算法选择个体规则如算法 1 所示。

**算法 1** 轮盘赌算法选择个体

输入: list pop, cumsum, ms。

输出: chrom list new\_pop。

```
begin:
    val new_pop = List < chrom >;
    val k = 0;
    for (i = 0; i < ms.size(); i++) {
        if (ms.get(i) < cumsum.get(k)) {
            chrom = pop.get(i);
            new_pop.add(chrom);
        }
        else {
            k++;
        }
    }
    return new_pop
end
```

(6) 染色体基因交叉。根据自然界的遗传规律,在遗传过程中,个体之间会两两交叉基因片段。互相交换基因片段的过程会产生新的可能解组合,即有可能产生新的随机森林 MCI 诊断模型参数组合。

例如,此时种群中有染色体  $i$  和  $j$  分别为 1011|0100|101|010 和 1101|1010|110|110, 两者在进行基因片段交叉时,若随机生成的交叉点为 7, 则重组规则为  $n_{ew_i} = i[1:6] + j[7:14]$ ,  $n_{ew_j} = j[1:6] + i[7:14]$ , 生

成新的染色体 1011|0110|110|110 和 1101|1000|101|010, 即对应产生新的随机森林 MCI 诊断模型参数组。图 4 为染色体基因片段交叉示例。

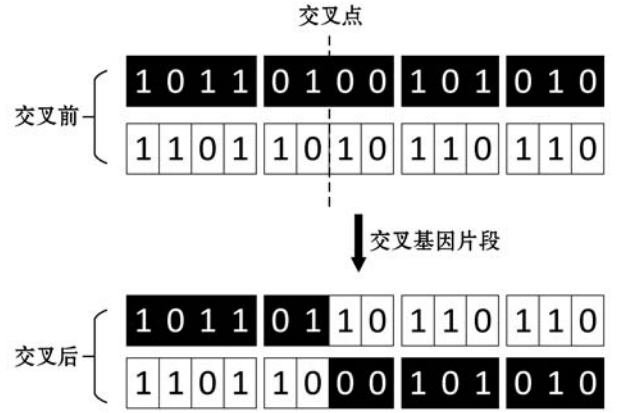


图 4 两个染色体基因片段交叉示例

种群中交叉遗传过程为:首先初始化交叉因子  $p_c$ , 再为每个随机森林参数组合个体  $i$  随机生成一个随机数作为突变概率  $c_{random_i}$ , 当  $c_{random_i} < p_c$  时,在种群中随机挑选一个个体与该个体进行基因片段交换。基因交叉过程如算法 2 所示。

**算法 2** 种群中基因交叉过程

输入: chrom list pop。

输出: chrom list new\_pop。

```
begin:
    val chrom_num = pop.size();
    val chrom_len = pop.get(0).size();
    for (i = 0; i < chrom_num; i++) {
        c_random_i = random(0, 1);
        if (c_random_i < pc) {
            mother_index = randint(0, chrom_num);
            index = randint(0, chrom_len);
            father = pop.get(i);
            mother = pop.get(mother_index);
            chrom = concat(father[0: index],
                           mother[index: chrom_len]);
        }
        else {
            chrom = pop.get(i);
        }
        new_pop.add(chrom);
    }
    return new_pop
end
```

(7) 基因变异。种群在遗传过程中,可能发生小概率的基因变异事件,加大了种群的多样性,为遗传创造了原材料<sup>[18]</sup>,通过此过程也有可能产生更有利的随机森林 MCI 诊断模型的参数组。

基因变异过程为:首先初始化突变因子  $p_m$ ,再为每个参数组合个体  $i$  随机创建一个突变概率  $m_{\text{random}_i}$ ,  $m_{\text{random}_i} < p_m$  时,在个体  $i$  基因中随机抽取一位进行变异。若原基因位二进制数为 0,则突变为 1;原基因为 1,则突变为 0。基因突变如算法 3 所示。

### 算法 3 基因变异算法

输入: chrom list pop。

输出: chrom list pop。

begin:

```
val chrom_num = pop.size();
val chrom_len = pop.get(0).size();
for (i=0; i < chrom_num; i++) {
    m_random_i = random(0,1);
    if (m_random_i < pm) {
        chrom = pop.get(i);
        index = randint(0,chrom_len);
        if (chrom.get(index) == 1) {
            chrom[index] = 0;
        }
        else {
            chrom[index] = 1;
        }
    }
}
```

return pop

end

(8) 终止条件判断。循环(3) - (6)过程,当循环次数达到预定值 gene\_time 次后,终止循环。

## 3 实验与结果分析

本文实验所使用的软件环境为操作系统 Windows 10,采用 Python 3.7 作为编程环境,开发环境为 PyCharm,硬件环境为 CPU AMD Ryzen 7 4800U 1.80 GHz,内存大小为 16 GB。

本节包括实验数据、参数设置和实验结果分析。本文在 Github 上对实验代码和数据进行了开源 ([https://github.com/zheng-gogo/Ga\\_randomForest\\_predictMci](https://github.com/zheng-gogo/Ga_randomForest_predictMci))。

### 3.1 实验数据

本文的数据来自于专业医学团队提供的关于老年人身体、心理多方面的 MCI 判断信息,去掉缺省特征值大于 1/4 的数据后,大约有 3 200 条,其中有 21 项属性,可分为 6 类,分别是基本信息、身体状况、病史、生活经历、生活习惯和调查问卷信息。表 2 是数据特征分类列表。

表 2 数据特征分类列表

类别	特征
基本信息	ID、性别、年龄、身高、体重、腰臀比
身体状况	收缩压、舒张压、空腹血糖值
病史	高血压史、高血压年数、糖尿病史、糖尿病年数、卒中史
生活经历	文化教育年数、婚姻状况(未婚、离婚、已婚、丧偶)
生活习惯	吸烟、饮酒
调查问卷信息	BMI、MMSE、ADL、CESD

根据特征值进行评估,MCI 诊断结果分类为 3 种:1 表示健康,2 表示患 MCI,3 表示患痴呆。

### 3.2 参数设置

本文将  $c_{\text{chrom}} = \{n_{\text{estimators}}, m_{\text{ax\_depth}}, m_{\text{ax\_features}}, m_{\text{in\_samples\_leaf}}\}$  作为一个参数组合构建染色体。随机森林 MCI 诊断模型参数对应基因位数、参数取值范围、参数实值与二进制基因数值的转化规则如表 3 所示。

表 3 遗传算法中随机森林 MCI 判定模型参数设置

参数名	基因位数	参数取值范围	转化规则
$n_{\text{estimators}}$	4	range(1, 121)	$n_{\text{estimators}} = 1 + B_{n_{\text{estimators}}} \times 8$
$m_{\text{ax\_depth}}$	4	range(1, 16)	$m_{\text{ax\_depth}} = 1 + B_{m_{\text{ax\_depth}}}$
$m_{\text{ax\_features}}$	3	range(1, 22)	$m_{\text{ax\_features}} = 1 + B_{m_{\text{ax\_features}}} \times 3$
$m_{\text{in\_samples\_leaf}}$	3	range(1, 8)	$m_{\text{in\_sample\_leaf}} = 1 + B_{m_{\text{in\_samples\_leaf}}}$

由表 3 可知,个体的基因数目  $c_{\text{chrom\_length}} = 14$ ,第一代种群 pop(0)随机生成,遗传算法参数设置如表 4 所示。

表 4 遗传算法初始值设置

参数名	说明	取值
$g_{\text{ene\_time}}$	遗传次数	100
$c_{\text{chrom\_num}}$	种群个体数目	10
$c_{\text{chrom\_length}}$	染色体长度	14
$p_c$	交叉因子	0.6
$p_m$	突变因子	0.01

### 3.3 实验结果分析

#### 3.3.1 遗传算法迭代过程分析

经过遗传算法,迭代 100 代,种群中随机森林 MCI 诊断模型参数最高准确率变化曲线如图 5 所示。

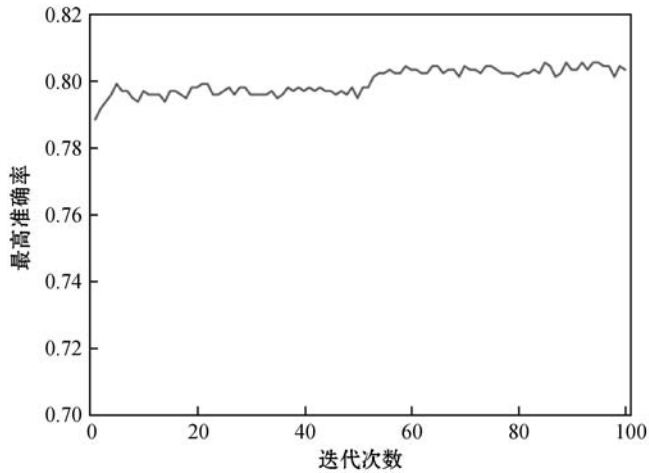


图 5 种群中随机森林 MCI 诊断模型参数最高准确率曲线

结合本文遗传算法分析图 5 可知:(1) 初始化种群表现不佳,但在 5 代左右,大概率的基因片段交叉快速产生了有利基因组合,最高准确率产生了第一次明显提升;但因轮盘赌算法保留个体的随机性,并未及时保留最优个体,最高准确率小幅回落。(2) 因除本文所述主要参数外,剩余对随机森林 MCI 诊断模型影响较小的参数的随机匹配,导致相同最优个体诊断效果也产生小幅度上下波动,但总体波动不大。(3) 在 50 代左右的演化过程中,由于基因变异产生了新的有利基因,出现第二次明显的准确率的提升。(4) 虽因其他参数的随机性,最高准确率在一定范围内有小幅度波动,但基本在 0.8 左右,总体趋于稳定,即种群收敛取得最优参数。

总体来说,遗传算法求取随机森林 MCI 预判模型最优参数过程中,种群中最高准确率呈上升趋势。

### 3.3.2 遗传算法和网格搜索算法对比

对于相同的取值范围,求取随机森林 MCI 诊断模型最优参数,遗传算法与网格搜索结果对比如表 5 所示。

表 5 遗传算法与网格搜索求取随机森林最优参数结果对比

算法	时间/s	最优参数组	准确率
网格搜索	8 758.88	(93,11,19,4)	0.803 4
遗传算法	196.45	(89,7,19,7)	0.802 1

可以看出,在时间性能上,网格搜索所用时间约是遗传算法的 45 倍,虽然得出的最优参数结果不同,但差距并不大,且二者对应的 MCI 诊断准确率差值也只有 0.001 3 左右。因此,综合比较分析,在本次实验过程中,遗传算法相较于网格搜索更有利于求随机森林的最优参数。

### 3.3.3 MCI 预判准确率分析

根据遗传算法求取到随机森林 MCI 预判模型最优参数组合为  $c_{chrom} = (n_{estimators} = 89, m_{ax\_depth} = 7, m_{ax\_features} =$

19,  $m_{in\_samples\_leaf} = 7)$ ,对同样的 MCI 样本数据进行诊断,随机森林与医学 MCI 诊断结果 F1 值对比如表 6 所示。

表 6 随机森林与医学 MCI 诊断结果 F1 值对比

诊断方式	正常	MCI	痴呆
随机森林诊断	0.887 3	0.389 3	0.847 8
医学人工诊断	0.845 2	0.374 7	0.823 5

可以看出:本文所采用的随机森林 MCI 诊断模型相比于医学人工诊断方式,对于正常、MCI 和痴呆三类的分类,F1 值这个综合评判指标都更高。因此,本文采用的随机森林模型配合遗传算法取得的最优参数对于 MCI 的诊断更为适用。

### 3.3.4 特征重要性分析

根据本数据随机森林输出 MCI 诊断过程特征的重要性排序如图 6 所示。

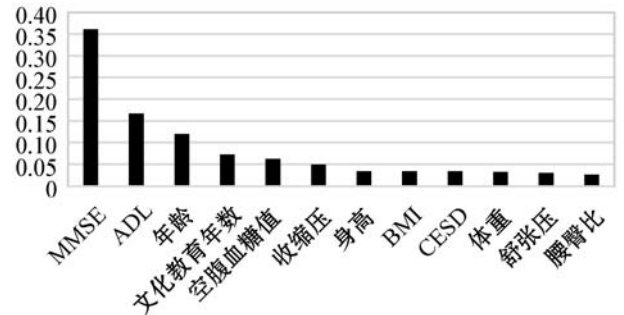


图 6 MCI 诊断过程特征重要性排序

由图 6 中的特征重要性排序可以看出以下几点:

(1) MMSE、ADL 和文化教育年数等特征在随机森林 MCI 诊断模型中起着至关重要的作用,特别是 MMSE 特征值占据着主导地位,与临床经验知识判定准则相符。

(2) 例如年龄、空腹血糖值和收缩压等因素,传统的判定上是较为粗略的经验判断,本文模型对重要性的评估验证了经验判断的合理性,且能依据本文模型更加准确地诊断 MCI。

(3) 本文基于现有检验指标,虽然没有重大的新的医学发现,但可从中获得启发,例如:在进行 MCI 诊断时,可将 50 岁以上的年龄进一步划分,对于不同的年龄段应用不同的诊断规则,更加准确地诊断 MCI。

## 4 结 语

本文提出基于随机森林对 MCI 进行自动诊断,并且通过遗传算法得出随机森林的最优参数。相比较于医学的诊断方式,本文诊断效果有了显著的提升,准确率提升了 5% 左右;同时采用遗传算法极大地缩短了

求取随机森林最优参数的时间。

传统的临床判断虽然包括了体检和问诊时的各项指标,但主要是依赖医师经验,无法给出不同因子的定量分析,本文对于特征重要性的评估给出了不同因素对诊断 MCI 的定量分析,在符合传统判定经验的同时能更准确地做出判断。

总之,通过以上机器智能化诊断方式减轻了医生的工作负荷,有利于辅助经验欠缺的医师做出临床决策,也方便老年人进行阿尔茨海默病自我预测与提前预防。

本文的未来工作主要有两个方面:第一,本文模型受限于既有数据中的各项指标,在加入更多因子(如基因组学研究)以后,模型有可能挖掘出更多诊断依据和新的病理机制;第二,虽然本文方式已经可以高效地进行 MCI 的自动诊断,但在遗传算法求取随机森林最优参数的过程仍存在需要改进的地方,如 1) 采用轮盘赌算法进行的个体选择的过程,由于随机性较大,致后代中适应度大的个体出现次数过多,使算法陷入局部最优<sup>[19]</sup>,并在局部最优解遗传多次,产生不必要的消耗;2) 经多次实验发现,在发生突变时产生的有利基因易打破局部最优解的僵局。因此,在未来的研究方向中,希望找到一种更加合理的方式进行个体选择,同时有效利用基因突变,提升算法性能。

## 参 考 文 献

- [ 1 ] Brookmeyer R, Johnson E, Ziegler Z, et al. Forecasting the global burden of Alzheimer's disease [ J ]. *Alzheimers Dementia*, 2007, 3(3) : 186 - 191.
- [ 2 ] 马菲. 太原市社区老年人轻度认知功能障碍现患及转归流行病学研究 [ D ]. 太原: 山西医科大学, 2009.
- [ 3 ] Misra C, Fan Y, Davatzikos C. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: Results from ADNI [ J ]. *NeuroImage*, 2009, 44(4) : 1415 - 1422.
- [ 4 ] Nasreddine Z, Phillips N, Bédirian V, et al. The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment [ J ]. *Journal of the American Geriatrics Society*, 2005, 53(4) : 695 - 699.
- [ 5 ] Chen K, Xu Y, Chu A, et al. Validation of the Chinese version of Montreal cognitive assessment basic for screening mild cognitive impairment [ J ]. *Journal of the American Geriatrics Society*, 2016, 64(12) : 285 - 290.
- [ 6 ] 温博文, 董文瀚, 解武杰, 等. 基于改进网格搜索算法的随机森林参数优化 [ J ]. *计算机工程与应用*, 2018, 54(10) : 154 - 157.
- [ 7 ] Lamini C, Benhlila S, Elbekri A. Genetic algorithm based approach for autonomous mobile robot path planning [ J ]. *Procedia Computer Science*, 2018, 127 : 180 - 189.
- [ 8 ] Breiman L. Random forest [ J ]. *Machine Learning*, 2001, 45 : 5 - 32.
- [ 9 ] Dudoit S, Fridlyand J. Bagging to improve the accuracy of a clustering procedure [ J ]. *Bioinformatics*, 2003, 19(9) : 1090 - 1099.
- [ 10 ] 周志华. 机器学习 [ M ]. 北京: 清华大学出版社, 2016 : 172.
- [ 11 ] Liaw A, Wiener M. Classification and regression by RandomForest [ J ]. *R News*, 2002, 2(3) : 18 - 22.
- [ 12 ] 王士霞. 基于决策树的在线学习行为分析 [ J ]. *河南科技学院学报(自然科学版)*, 2015, 43(5) : 67 - 70, 75.
- [ 13 ] Ming D, Zhou T, Wang M, et al. Land cover classification using random forest with genetic algorithm-based parameter optimization [ J ]. *Journal of Applied Remote Sensing*, 2016, 10(3) : 035021.
- [ 14 ] 黄辉先, 史忠科. 利用遗传算法搜索全局最优的一种混合算法 [ J ]. *电子与信息学报*, 2001(9) : 875 - 878.
- [ 15 ] 陈禹, 毛莺池. 基于随机森林和遗传算法的 Ceph 参数自动调优 [ J ]. *计算机应用*, 2020, 40(2) : 347 - 351.
- [ 16 ] Goldberg D. Genetic algorithm in search optimization and machine learning [ M ]. Addison Wesley Longman Publishing, 1989.
- [ 17 ] 夏桂梅, 曾建潮. 一种基于轮盘赌选择遗传算法的随机微粒群算法 [ J ]. *计算机工程与科学*, 2007(6) : 51 - 54.
- [ 18 ] 王涛. 遗传算法及其应用 [ J ]. *新乡学院学报*, 2008, 25(1) : 56 - 58.
- [ 19 ] 易欣, 郭武士, 赵丽. 利用自适应选择算子结合遗传算法的机器人路径规划方法 [ J ]. *计算机应用研究*, 2020, 37(6) : 1745 - 1749.

## (上接第 222 页)

- [ 18 ] Zhang K, Sun M, Han T X, et al. Residual networks of residual networks: Multilevel residual networks [ J ]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, 28(6) : 1303 - 1314.
- [ 19 ] Zhang X, Chen M H, Qin Y. NLP-QA framework based on LSTM-RNN [ C ] // 2nd International Conference on Data Science and Business Analytics, 2018 : 307 - 311.
- [ 20 ] Chung E, Park J G. Sentence-chain based Seq2seq model for corpus expansion [ J ]. *ETRI Journal*, 2017, 39(4) : 455 - 466.
- [ 21 ] Wu Y, Wei F R, Huang S H, et al. Response generation by context-aware prototype editing [ C ] // 33rd AAAI Conference on Artificial Intelligence, 2018 : 7281 - 7288.
- [ 22 ] Guo D, Zhou W G, Li A Y, et al. Hierarchical recurrent deep fusion using adaptive clip summarization for sign language translation [ J ]. *IEEE Transactions on Image Processing*, 2019, 29(99) : 1575 - 159.