

# 基于随机森林的帕金森疾病诊断模型构建研究

王小科, 晏峻峰\*

(湖南中医药大学 信息科学与工程学院, 湖南 长沙 410208)

**摘要:**近些年来,根据帕金森疾病(PD)患者的语音数据对该疾病做出诊断成为一种行之有效的疾病诊断方法。首先,针对语音数据集中存在非均衡数据和噪声样本的问题使用SVM SMOTE过采样技术,利用支持向量机分类器寻找支持向量并在此基础上合成新的样本以达到均衡数据集的目的;为了减少数据维度,降低学习难度,运用信息增益特征选择对所有特征属性计算数值并划分数据集以此来获得信息增益,根据信息增益的大小排序选取得到八个特征作为最优特征组合;最后,构建随机森林帕金森疾病诊断模型,并采用网格搜索和交叉验证相结合的方式进行参数调优,进一步优化模型,实现诊断模型准确率的进一步提高。实验结果表明,优化后模型的准确率、灵敏度和特异度均有提升,分别为96.59%、94.81%和95.49%,且准确率均高于支持向量机、最邻近节点算法、朴素贝叶斯和决策树等决策模型。

**关键词:**SVM SMOTE;信息增益;随机森林;网格搜索;交叉验证

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2023)04-0154-07

doi:10.3969/j.issn.1673-629X.2023.04.023

## Research on Construction of Parkinson's Disease Diagnosis Model Based on Random Forest

WANG Xiao-ke, YAN Jun-feng\*

(School of Information Science and Engineering, Hunan University of Chinese Medicine, Changsha 410208, China)

**Abstract:**In recent years, the diagnosis of Parkinson's disease based on the speech data of patients with Parkinson's disease has become an effective disease diagnosis method. Firstly, we use the SVM SMOTE oversampling technique to solve the problem of unbalanced data and noise samples in the speech data set, and use the SVM classifier to find the support vector and synthesize new samples on this basis to achieve the purpose of the balanced data set. In order to reduce the data dimension and the difficulty of learning, we use the information gain feature selection to calculate the value of all feature attributes and divide the data set to obtain the information gain, and select eight features as the optimal feature combination according to the size of the information gain. Finally, a random forest Parkinson's disease diagnosis model is constructed, and the parameters are optimized by the combination of grid search and cross-validation to further optimize the model, so as to further improve the accuracy of the diagnosis model. The experimental results show that the accuracy, sensitivity and specificity of the optimized model are improved, which are 96.59%, 94.81% and 95.49%, respectively, and the accuracy is higher than that of support vector machine, nearest neighbor algorithm, naive Bayes and decision trees.

**Key words:**SVM SMOTE; information gain; random forest; grid search; cross-validation

### 0 引言

帕金森疾病(Parkinson's disease, PD)为慢性进展性运动障碍病症,可受诸多复杂因素导致锥体外系功能性障碍,且随着人类寿命的延长,PD患者的数量也在不断增加<sup>[1]</sup>。PD具有持续静止性颤动、肌肉强直、运动障碍等临床表现,同时还伴有语言和发声受限的问题<sup>[1-2]</sup>。目前新型的帕金森诊断方法包括:基于手绘图特征诊断帕金森病、步态分析和面部表情分析诊

断PD,这些研究取得了一定的成果,但是也存在检测过程复杂,对病人身体产生物理损伤等问题,不易进行大范围的PD检测。近年来,大量研究学者提出了各种非侵入性的方法来检测帕金森病的症状。其中在PD的病症表现中,发现几乎所有的PD患者会出现并发症引起的声带损伤,从而出现某种程度的语音障碍<sup>[3]</sup>,因此,检测语言障碍是实现高准确率的PD疾病诊断的可靠手段。

收稿日期:2022-06-06

修回日期:2022-10-08

基金项目:2021年湖南省教育科学研究重点项目(21A0250);湖南中医药大学中医学一流学科开放基金(2022ZYX08)

作者简介:王小科(1999-),女,硕士研究生,研究方向为机器学习、数据挖掘;通讯作者:晏峻峰(1965-),女,教授,博导,研究方向为人工智能及其应用。

该文采用 UCI 公开的 Speech 帕金森语音数据集,首先针对数据集中存在噪声样本以及非均衡数据的问题,使用 SVM SMOTE 过采样技术均衡数据集,接着采取信息增益特征选择数据集中的八个特征作为最优特征组合并构建 RF 模型。为进一步提升算法的性能,使用网格搜索与交叉验证相结合的方式寻找 RF 最优参数,进一步提高了模型的分类准确率。

## 1 相关研究

近年来在基于语音信息的 PD 诊断方面的研究,首先是对语音数据提取特征,然后采用机器学习建模的方式进行分类。B. E. Saker 等人设计了一个用于数据收集的计算机辅助系统,将参与实验人员的发声数据收集出来,并设计合理的实验方法,从中分析提取出具有较大 PD 信息的属性特征。Little 等人<sup>[4]</sup>对语言障碍的检测则是通过提取其中的非线性特征,进一步对患者病情的发展情况做出判断。目前国内外从事基于语音信息 PD 诊断的学者,数据大部分来自这两位研究人员。在特征提取和特征压缩方面,主要研究方法为:主成分分析(Principal Component Analysis, PCA<sup>[5]</sup>)、穷举法<sup>[6]</sup>、基于局部学习的特征选择法<sup>[6]</sup>、启发式特征选择等<sup>[6]</sup>。在疾病诊断模型方面,主要研究方法为:最邻近节点算法(K-Nearest Neighbor, KNN<sup>[6]</sup>)、贝叶斯网络(Naive Bayes, NB<sup>[7]</sup>)、随机森林(Random Forest, RF<sup>[7]</sup>)、最小均方误差回归<sup>[8]</sup>、BP 神经网络<sup>[8]</sup>、支持向量机(Support Vector Machine, SVM<sup>[5,9]</sup>)等。Kursun 等<sup>[10]</sup>通过基于信息的方法进行排列分析,以得到最佳数据,文献中采用皮尔逊相关系数方法来优选数据;Chen<sup>[11]</sup>等人使用 PCA 降维方法,然后再采用模糊 K 近邻方法进行重新采样的估计方法;通过生物启发算法,如 Olivares 等基于蝙蝠算法设计的生物极限学习机可以进行分类试验,但是,该方法只能采用梅尔顿倒谱系数,并没有考虑到数据遗漏的问题;李勇明等提出了基于语音样本重复剪辑和随机森林的帕金森诊断算法对帕金森数据样本进行分类诊断,达到了较好的分类准确度,但该方法语音特征较多,效率低下;张琼等人<sup>[12]</sup>采用改进 PSO-SVM 算法对不同性能的粒子动态分配惯性权重和学习因子,提高支持向量机模型的学习能力和泛化能力,提高了对疾病的识别精度;郭东伟<sup>[13]</sup>采用样本约简算法和 3 种特征降维算法分别结合 3 种机器学习分类算法对样本分类,得出 ELM+NNIR+KPCA 的最佳组合模型;谭言丹<sup>[14]</sup>设计基于 AdaBoost 的特征筛选方法并开发基于正则化损失函数的 XGBoost 来实现最终病情诊断;罗正潮<sup>[15]</sup>通过中文语音识别 PD 患者情况,通过发声分析、韵律分析和轮替运动分析提取出患者身体状况

的语音特征,使用 XGBoost 算法提高了中文 PD 识别的优势;马超等人<sup>[16]</sup>利用混沌理论和高斯变异方法改进樽海鞘算法,有效地解决了模型的参数设定和最优特征选择问题;张小恒等人<sup>[17]</sup>提出两步式稀疏迁移学习算法;采用语音段特征同时优选的快速卷积稀疏编码算法和采用联合局部结构信息分布对齐算法,在保持各自样本结构信息的同时最小化分布误差,从而显著提高了算法的准确率。

## 2 研究方法

### 2.1 方法总述

算法整体流程如图 1 所示。

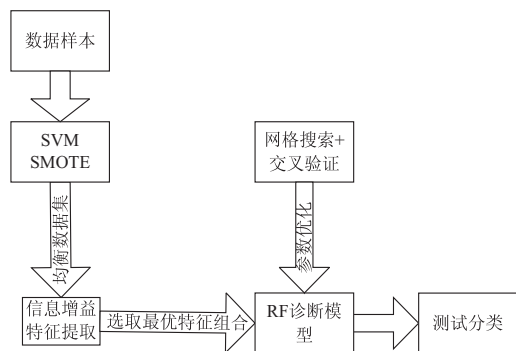


图1 算法整体流程

由于该文所使用的数据集存在非均衡数据和样本噪声的问题,使用 SVM SMOTE 过采样技术在没有引入更多数据的情况下平衡二分类数据;使用信息增益特征提取根据信息增益大小排序选取得到 8 个特征作为最优特征组合;构建 RF 帕金森疾病诊断模型,并使用网格搜索和交叉验证相结合的方式进行参数调优;最后使用该模型对测试集进行样本分类。

### 2.2 SVM SOMTE

SVM SOMTE 算法使用一个 SVM 分类器寻找支持向量,然后在支持向量上合成新的样本。SVM SMOTE 侧重于沿决策边界增加少数点,这背后的论点是,围绕该边界的实例对于估计最佳决策边界至关重要。算法思路如下:包含边界信息的原始支持向量机通过从不平衡数据集中学习的参数 SVM 粗略识别  $D$ 。在初始过程中,通过交叉验证选择适当的内核和超参数  $C$ ,其中 G-means 被选为最佳指标。然后将属于少数的原始支持向量标记为  $SV_0 = \{X_{SV_i}^{(0)}\}$ ,并通过删除  $SV_0$  构建一个用于进一步引导的新数据集  $D \sim = D \setminus SV_0$ 。依次在  $D \sim$  上执行自举,每个采样结果的规模为  $|D \sim|$ 。此外,Extrapolation Borderline-SMOTE 执行  $D \sim$  和  $SV_0$  的聚合数据集。之后,具有新合成样本的合并数据集用于元 SVM 的训练,未采样的原始数据作为调整参数的验证集。最后,  $K$ ,  $SVMS$  以相同的权重聚合以形成集成分类器。具体来说,模型中的

默认参数最初设置为  $\alpha = 0.5$ ,  $K = 100$ , 该算法主要流程如下:

(1) 输入整个数据集, SVMs -  $K$  的数量, Extrapolation Borderline-SMOTE 的过采样率  $\alpha\%$ 。

(2) 在原始数据集上训练, 通过在交叉验证中选择一个内核和超参数  $C$  来拟合软边距 SVM 并识别支持向量,  $SV_0 = \{X_{SV_0}^{(0)}\}$  属于少数。

(3) 将  $t$  从 1 到  $K$  循环。

(4) Bootstrap 在  $D \sim D \setminus SV_0$  上获得的采样结果为  $D_t^-$  和  $D_t^c$ , 未依次采样  $t$ 。

(5) 得到并集为  $D_t^- \cup SV_0$ , 执行带有采样率为  $\alpha\%$  的 extrapolation borderline-SMOTE。

(6)  $D_t^- \cup SV_0$  和合成样本联合作为训练集  $D_t$ , 通过带有在  $D_t^c$  上验证表现的超参数获得软边距 SVM  $f_t(x)$ , 通过在  $D_t^c$  上验证表现的超参数。

(7) 输出支持向量机的集合  $F(x) = \text{sgn}(\sum f_t(x))$ 。

### 2.3 信息增益特征选择

信息增益通过计算信息熵而来, 量化了信息去除不确定性的程度, 可以通过信息增益的大小为变量排序进行特征选择。信息量与概率两者之间呈单调递减关系, 概率越小, 信息量越大。

#### 2.3.1 信息量

$$I(u_i) = \log_2 \frac{1}{P(u_i)} = -\log_2 P(u_i)$$

其中,  $u$  表示发送的消息,  $u_i$  表示发送信息  $u$  的一种类型。

#### 2.3.2 先验熵

信息熵表示信息量的数学期望, 是信源发出信息前的平均不确定性, 也称为先验熵。信息熵的定义如下所示:

$$\begin{aligned} \text{Ent}(U) &= \sum_{i=1}^k P(u_i) I(u_i) = \sum_{i=1}^k P(u_i) \log_2 \frac{1}{P(u_i)} = \\ &= - \sum_{i=1}^k P(u_i) \log_2 P(u_i) \end{aligned}$$

当  $u_i$  中某一种类型的概率为 1 时, 即没有发送消息的不确定性, 信息熵  $\text{Ent}(U) = 0$ ;

当  $u_i$  对应的概率相同时, 即概率都为  $1/k$ , 信息熵  $\text{Ent}(U) = \log_2 k$ 。

#### 2.3.3 后验熵

信息熵  $\text{Ent}(U)$  表示在发出信息  $U$  之前存在的不确定性, 在接收搭配信息  $V$  之后, 信息  $U$  的不确定性会发生改变, 即后验熵, 它是接收到一定的信息后, 对信息  $U$  进行的后验判断, 定义如下:

$$\text{Ent}(U|v_j) = \sum_{i=1}^k P(u_i|v_j) \log_2 \frac{1}{P(u_i|v_j)}$$

$$- \sum_{i=1}^k P(u_i|v_j) \log_2 P(u_i|v_j)$$

考虑所有信息  $V$  时, 得到的后验熵的期望 (又称条件熵) 如下,  $\text{Ent}(U|V)$  是指在  $V$  结束之后存在对信息  $U$  的平均不确定性, 通常由随机干扰引起。

$$\begin{aligned} \text{Ent}(U|V) &= \sum_{j=1}^m P(v_j) * \text{Ent}(U|v_j) = \\ &= \sum_{j=1}^m P(v_j) \left( - \sum_{i=1}^k P(u_i|v_j) \log_2 P(u_i|v_j) \right) \end{aligned}$$

#### 2.3.4 信息增益

信息增益是指先验熵与后验熵差值部分, 表示信息消除不确定性的程度, 定义如下:

$$\text{Gains}(U, V) = \text{Ent}(U) - \text{Ent}(U|V)$$

特征选择原理: 对数据集选择特征时, 以目标标量作为信息  $U$ , 由特征变量作为信息  $V$ , 代入公式计算信息增益, 以信息增益的值按照大小顺序来确定特征的顺序, 以此进行特征选择。信息增益越大, 表示变量消除不确定性的能力越强。

### 2.4 随机森林

随机森林 (Random Forest, RF) 主要思想是在原始的数据集上随机挑选  $P$  个样本作为训练集以生成相应数量的决策树, 在挑选最佳属性时, RF 并非像决策树一样让所有的属性都参与选择, 而是随机选择  $Q$  个属性用于构建决策树, 然后将这些决策树构成一个森林, RF 中每一棵决策树之间是独立无关联的。当 RF 构建完成之后, 输入新的数据样本, 决策森林中的所有决策树均会做出判断并将输出最多的类别作为最终的模型诊断类别。

具体实现过程如下:

(1) 对原始训练集  $M$  使用 bootstrap 方法, 在  $M$  中有放回随机选取  $X$  个样本集合, 构成  $X$  个分类树, 每次没有被选中的数据样本构成  $K$  个袋外数据;

(2) 假设有  $ma$  个变量, 在决策树的所有节点处均随机抽取  $mt$  变量 ( $mt, n, ma$ ), 然后在  $mt$  挑选分类能力最强的变量, 最后通过测量每一个分类点确定变量分类的阈值;

(3) 所有决策树均可以得到最大限度的生长, 也无需任何修剪;

(4) 将生成的所有分类树构成 RF 并输入新的数据进行分类判断, 根据树分类器的投票结果作为输出类别的判别。

## 3 实验过程及分析

实验总共包括三个部分: (1) 针对数据集中存在噪声样本以及非均衡数据的问题, 使用 SVM SMOTE 过采样技术对数据进行均衡处理。 (2) 使用信息增益



特征提取算法对该数据集进行特征提取,选取八个特征作为最优特征组合,将样本映射到更加低维的特征空间,减少了系统的计算时间,同时提高了模型的准确率。(3)构建 RF 帕金森疾病诊断模型,由于 RF 参数  $n\_estimators$  的取值范围过大,首先使用学习曲线的方式学习参数  $n\_estimators$  的大致范围,接着使用网格搜索和交叉验证相结合的方式寻找 RF 模型的最佳参数组合,进一步提升了诊断模型的准确率。

### 3.1 数据集及预处理

#### 3.1.1 实验数据

该数据集是由英国牛津大学的 Max Little<sup>[4]</sup>与科罗拉多州丹佛市的国家语音和语音中心联合建立。由三十一个患者发送固定的元音‘a’,其中二十三人属于 PD 患者,八人健康,从而得到一组生物医学语音检测信息,总计 197 条信息,23 个属性特征以及 1 个标签。属性分别为:平均声部基频 MDVP:Fo (Hz),最大声部基频 MDVP:Fhi (Hz),最小声部基频 MDVP:Flo (Hz),基频变化的几个度量 MDVP:Jitter (%),MDVP:Jitter (Abs)、MDVP:RAP、MDVP:PPQ、Jitter:DDP,振幅变化的几个度量 MDVP:Shimmer、MDVP:Shimmer (dB)、Shimmer:APQ3、Shimmer:APQ5、MDVP:APQ、Shimmer:DDA,语音状态中噪声与音调分量之比的两种测量 NH、HNR,两种非线性动态复杂性测量 RPDE、D2,信号分形缩放指数 DFA,基频变化的三种非线性测量 spread1、spread2、PPE,姓名 name。标签为 status,1 表示患者,0 表示健康。

#### 3.1.2 数据预处理

在该数据集中姓名 name 属性列描述患者的姓名,对实验没有实质的用途和意义,所以在数据集中将姓名属性列移出,即所使用的数据集为 22 个属性列,1 个标签列。另外,由于数据集的特征取值范围各不相同,该文将数据做归一化处理,将数据映射到 0-1 之间,使得特征之间持有相同的度量尺度。公式如下:

$$x_{scale} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

其中,  $x_{min}$  和  $x_{max}$  是指要归一化数据的最小边界和最大边界。 $x$  为原始数值,  $x_{scale}$  为归一化处理之后的数值。

#### 3.1.3 SVM SMOTE 均衡数据集

该语音数据样本标签严重不平衡,其中标签为 1 的样本数量为 147,标签为 0 的样本数量为 48。从数据角度来看,造成数据稀疏问题:数据的缺失导致模型对稀疏数据刻画能力不足,无法对此类样本进行分类;从分类器的角度看,会造成决策边界偏移问题:因为多数类样本数量远大于少数类样本数量,导致学习的分类界限更偏向于多数类,造成决策边界的偏移,最终使得模型的分类性能下降。SVM SMOTE 算法的核心思想是聚合各种支持向量机,通过向正确的方向构建合成少数样本来修正初始决策边界,再以插值的方式为少数样本生成合成样本以达到均衡数据集的目的。图 2 为数据集均衡前数据集各属性和标签的数据分布直方图,图 3 为数据集均衡后各属性和标签的数据分布直方图,两图中最后一个直方图为标签数据分布图。

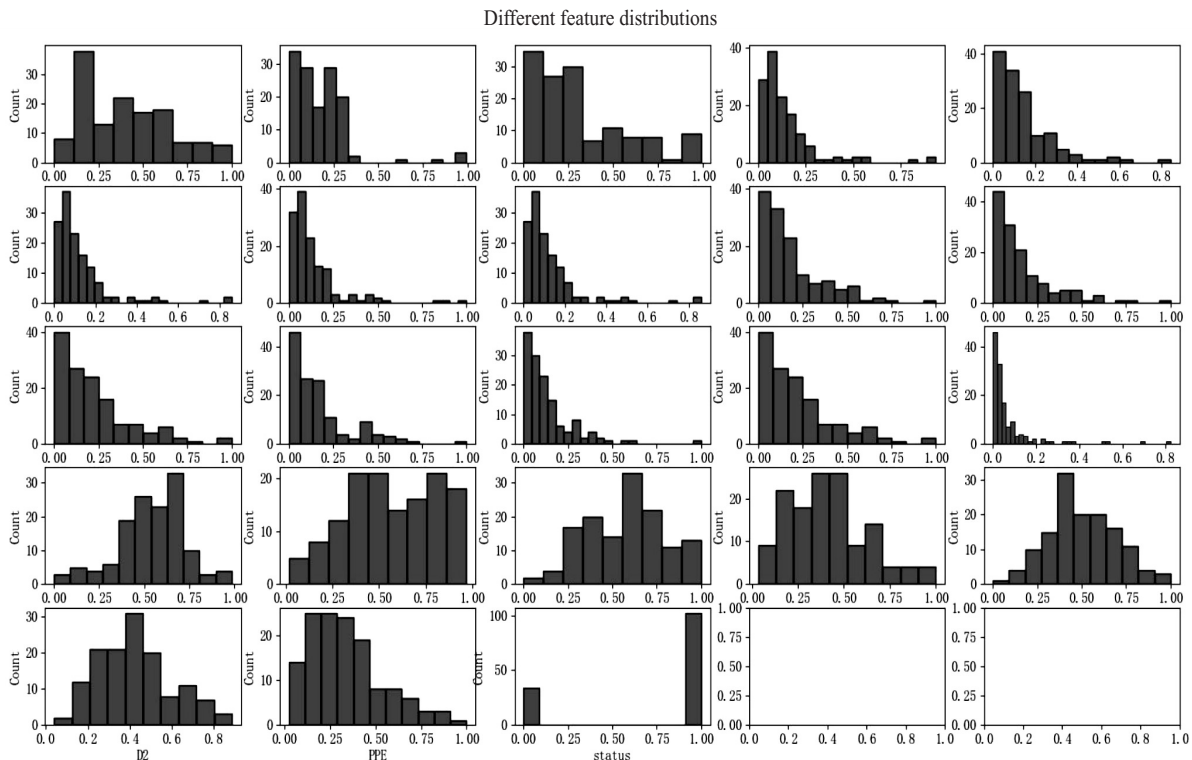


图2 SVM SMOTE 均衡前数据分布直方图

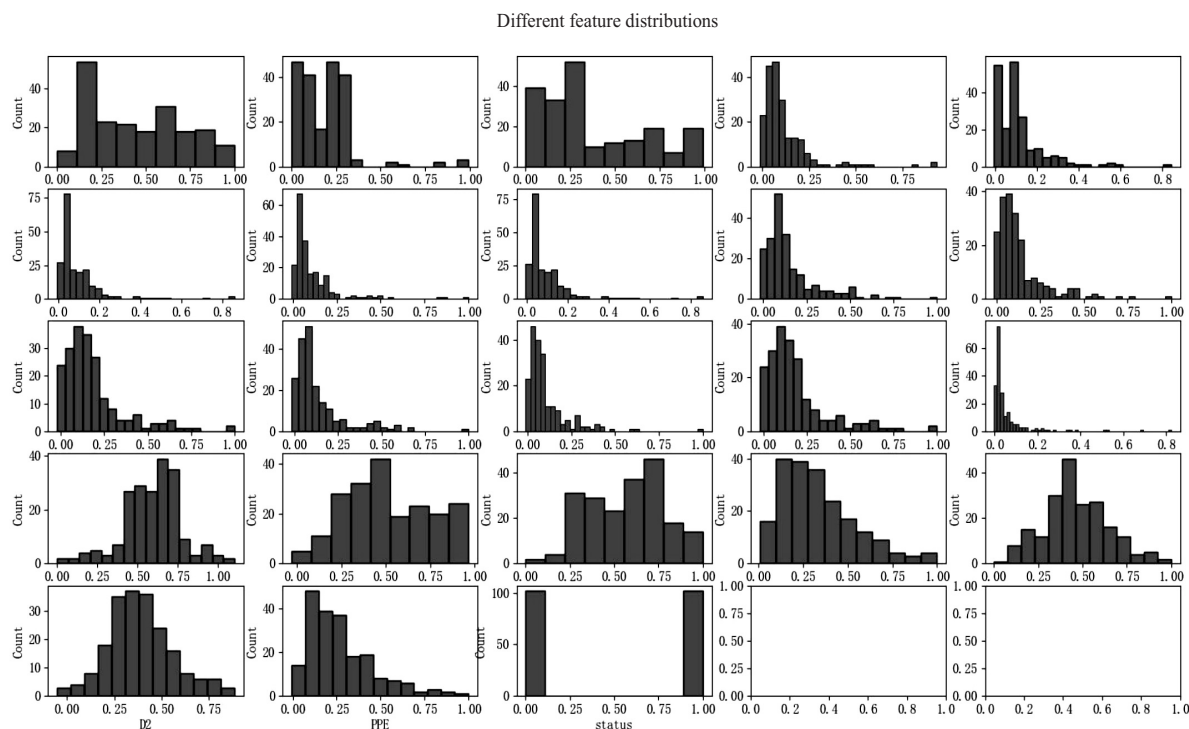


图 3 SVM SMOTE 均衡后数据分布直方图

### 3.2 实验环境与参数设置

#### 3.2.1 实验环境

PD 诊断模型运行在 windows10 系统 PyCharm2021.3.2 平台下,基于 Anaconda 开发环境下 Python3.8 编写运行。实验以 70% 的 PD 数据集作为训练集,30% 作为测试集,根据测试集的分类结果对模型进行评估。

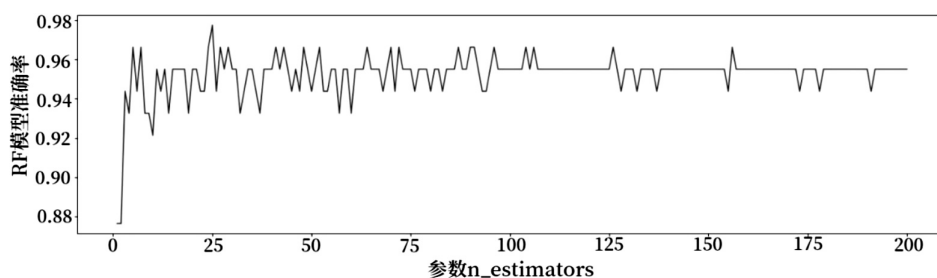
#### 3.2.2 信息增益特征选择选取最优特征组合

该文采用的语音数据集包含 196 个样本,22 个特征,1 个标签。与样本量相比,特征数量相对较多,大胆假设这些特征中必然存在无关特征和冗余特征,这些特征对算法毫无意义,无法提升算法的性能。另外,还存在一些冗余特征可以由其他特征推断而来,这种情况下,可以通过特征选择减少数据的特征维度,过滤数据噪声提高模型精度,同时降低学习任务的难度。信息增益特征选择通过比较特征值划分数据集所获取的信息增益,选取信息增益最高的特征划分数据集。具体流程为:(1)根据公式计算原始信息熵  $H$ ;(2)选

取一个特征,用特征值进行分类并计算类别的信息熵,以一定的比例加和得出该划分方式下的信息熵  $H'$ ;(3)计算信息增益,即  $H - H'$ ;(4)根据(2)、(3)计算所有特征属性对应的信息增益,保留信息增益较大的特征属性。经过算法选择后选取出的最优特征数量为 8,分别为 MDVP:F0(Hz), MDVP:Fhi(Hz), MDVP:Flo(Hz), Shimmer:DDA, HNR, spread2, PPE, D2。

#### 3.2.3 随机森林参数优化

构建 RF 模型,使用网格搜索<sup>[13-14]</sup>参数寻优寻找 RF 全局最优参数组合时,由于参数  $n\_estimators$  搜索的范围较大,倘若网格搜索遍历的步长较小且要包含最优解,那么会出现计算量过大而降低电脑运行速度等问题。通过刻画参数  $n\_estimators$  学习曲线的方式编码绘制 RF 参数的学习曲线,确定参数大致取值范围,再使用网格搜索在该范围内进行精细搜索,进一步获得包括参数  $n\_estimators$  在内的参数的最优组合。图 4 为参数  $n\_estimators$  的学习曲线,由图可知参数的最优值大致在 25 左右,因此确定网格搜索该参数的范

图 4 参数  $n\_estimators$  学习曲线

围为 20 到 30。对于其他参数由于在参数  $n\_estimators$  确定后的取值范围较小,可直接使用网格搜索寻找最佳值。经网格搜索计算,该参数的最优组合为  $n\_estimators = 25$ ,  $criteriion = "gini"$ 。

### 3.3 评价指标

使用准确率 (Accuracy)、灵敏度 (Sensitivity)、特异度 (Specificity) 作为算法的判断指标,以此来检验文中模型的合理性。准确率指数据中被正确诊断的病人数量与总样本数的比例;而灵敏度又叫真阳性比例,即实际发病且被准确诊断的病人所占比例;特异度又称为真阴性率,是指实际无病并能准确检测的病历所占比例。具体包括:

$$\text{准确率} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{灵敏度} = \frac{TP}{TP + FN}$$

$$\text{特异度} = \frac{TN}{TN + FN}$$

其中, TP 为真阳性例数, TN 为真阴性例数, FP 为假

阳性例数, FN 为假阴性例数。

### 3.4 结果分析

表 1 分别为经过 SVM SMOTE 算法均衡数据集、通过特征提取以及网格搜索交叉验证优化 RF 参数后的模型评估指标。由表可知,原始 RF 算法的准确率、灵敏度和特异度分别为 92.38%、96.54%、80.17%;经过 SVM SMOTE 算法均衡数据之后,准确率提升至 94.76%,灵敏度和特异度也分别提高至 94.37% 和 97.33%;特征提取获取最优参数组合后,模型的准确率、灵敏度和特异度分别为 96.02%、94.93% 和 95.42%;利用网格搜索寻得 RF 的最佳参数之后,RF 模型的准确率、灵敏度和特异度分别为 96.59%、94.81% 和 95.49%。

为了验证所提出方法的有效性和先进性,基于 SVM SMOTE 和信息增益提取目标特征子集之后构建的 RF 模型,就多项指标与现有研究中具有代表性的模型,如 SVM、KNN、NB、XGBoost、DT 进行了详细对比,表 2 展示了对比结果。

表 1 RF 诊断模型结果对比 %

算法	Accuracy	Sensitivity	Specificity
原始 RF 算法	92.38	96.54	80.17
SVM SMOTE 均衡后的 RF 算法	94.76	94.37	97.33
特征提取后的 RF 算法	96.02	94.93	95.42
优化后的 RF 诊断模型	96.59	94.81	95.49

表 2 不同分类器性能对比 %

分类器	Accuracy	Sensitivity	Specificity
RF	96.59	94.81	95.49
SVM	90.32	90.72	90.45
KNN	94.08	90.21	98.18
NB	87.03	90.41	82.39
XGBoost	95.24	90.87	91.24
DT	92.56	96.28	80.47

从表中可以看出,RF 模型实现了最佳的准确率,其灵敏度稍低于 DT,其特异度低于 KNN,然而当指标数值在 85% 到 95% 的范围时,即可判定该指标性能良好。可见,文中模型具备较好的泛化性能。

## 4 结束语

文中数据来源于 kaggle 官网,由英国牛津大学的 Max Little<sup>[6]</sup>与科罗拉多州丹佛市的国家语音和语音中心合作创建。通过使用 SVM SMOTE 均衡算法均衡数据集使得样本达到平衡,特征选择根据信息增益的大小排序选取得到 8 个特征,进而建立了基于 RF 的 PD 疾病诊断模型,并使用网格搜索和交叉验证相

结合的方式进行参数调优进一步提高模型的准确率。与传统 RF 相比,PD 诊断模型在准确率、灵敏度和特异度上的表现均有提高,实验的有效性得以证明。

### 参考文献:

- [1] 李勇明,杨刘洋,刘玉川,等. 基于语音样本重复剪辑和随机森林的帕金森病诊断算法研究[J]. 生物医学工程学杂志,2016,33(6):1053-1059.
- [2] 王宗宝,黄永志,张新静,等. 帕金森病患者局部场电位信号多频耦合特征分析[J]. 生物医学工程学杂志,2015,32(4):874-880.
- [3] 谢文宾. 基于语音特征学习的说话人确认与帕金森诊断[D]. 重庆:重庆大学,2016.

- [4] LITTLE M A, MCSHARRY P E, HUNTER E J, et al. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease[J]. IEEE Transactions on Bio-Medical Engineering, 2009, 56(4): 1015-1022.
- [5] HARIHARAN M, POLAT K, SINDHU R. A new hybrid intelligent system for accurate detection of Parkinson's disease[J]. Computer Methods & Programs in Biomedicine, 2014, 113(3): 904-913.
- [6] SAKAR B E, ISENKUL M M, SAKAR C O, et al. Collection and analysis of a parkinson speech dataset with multiple types of sound recordings[J]. IEEE Journal of Biomedical & Health Informatics, 2013, 17(4): 828-834.
- [7] PÉREZ C J, NARANJO L, MARTÍN J, et al. A latent variable-based Bayesian regression to address recording replications in Parkinson's disease[C]//22nd European signal processing conference. Lisbon: IEEE, 2014.
- [8] ISLAM M S, PARVEZ I, HAI D, et al. Performance comparison of heterogeneous classifiers for detection of Parkinson's disease using voice disorder (dysphonia)[C]//International conference on informatics. Shanghai: IEEE, 2014.
- [9] PRASHANTH R, ROY S D, MANDAL P K, et al. Parkinson's disease detection using olfactory loss and REM sleep disorder features[C]//2014 36th annual international conference of the IEEE engineering in medicine and biology society. Chicago: IEEE, 2014.
- [10] SAKAR C O, KURSUN O. Telediagnosis of Parkinson's disease using measurements of dysphonia. [J]. Journal of Medical Systems, 2010, 34(4): 591-599.
- [11] CHEN H L, HUANG C C, YU X G, et al. An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach[J]. Expert Systems with Applications, 2013, 40(1): 263-271.
- [12] 张 琼, 丁卫平, 景 炜, 等. 基于改进 PSO-SVM 算法的帕金森疾病诊断研究[J]. 计算机与数字工程, 2019, 47(8): 1851-1855.
- [13] 郭东伟. 基于语音特征的帕金森病诊断算法研究[D]. 重庆: 重庆大学, 2020.
- [14] 谭言丹. 基于语音信号和手绘图的帕金森病自动诊断[D]. 太原: 中北大学, 2020.
- [15] 罗正潮, 刘 蔚, 余 越, 等. 基于中文语音信号的帕金森疾病远程检测[C]//2021 中国自动化大会论文集. 北京: 出版者不详, 2021: 301-307.
- [16] 马 超, 谭 旭. 基于樽海鞘算法优化的帕金森病早期诊断模型研究与并行优化[J]. 计算机应用研究, 2021, 38(9): 2726-2731.
- [17] 张小恒, 张馨月, 李勇明, 等. 面向帕金森病语音诊断的非监督两步式卷积稀疏迁移学习算法[J]. 电子学报, 2022, 50(1): 177-184.
- .....
- (上接第 153 页)
- 法[J]. 计算机科学与探索, 2021, 15(6): 1155-1164.
- [15] MIRJALILI S, LEWIS A. The whale optimization algorithm[J]. Advances in Engineering Software, 2016, 95: 51-67.
- [16] 冯 茜, 李 擎, 全 威, 等. 多目标粒子群优化算法研究综述[J]. 工程科学学报, 2021, 43(6): 745-753.
- [17] 张晓凤, 王秀英. 灰狼优化算法研究综述[J]. 计算机科学, 2019, 46(3): 30-38.
- [18] 汪雪莹, 贺兴时. 飞蛾扑火优化算法的研究及改进[J]. 河南科学, 2021, 39(7): 1052-1061.
- [19] 唐延强, 李成海, 宋亚飞, 等. 自适应变异麻雀搜索优化算法[J]. 北京航空航天大学学报, 2022, 48(6): 1-6.
- [20] 聂方鑫, 王宇嘉. 基于自适应 t 分布与随机游走的麻雀搜索算法[J/OL]. 电子科技: 1-7 [2022-05-17]. DOI: 10.16180/j.cnki.issn1007-7820.2023.07.011.
- [21] 汤安迪, 韩 统, 徐登武, 等. 基于等级制度和布朗运动的混沌麻雀搜索算法[J]. 空军工程大学学报: 自然科学版, 2021, 22(3): 96-103.