

基于组合模型的新型镇静药物疗效分析与预测

康俊达, 汪波, 夏雅雯, 万优艳*

江汉大学人工智能学院, 湖北 武汉

收稿日期: 2023年12月12日; 录用日期: 2024年1月7日; 发布日期: 2024年1月15日

摘要

临床研究可验证科学假设、提供循证医学证据, 不仅是临床诊治方法及医学学科发展的重要保证, 还是新药物研究中的关键环节。本文对手术中的一种原有镇静药物与一种新型镇静药物的临床试验数据进行研究, 分析新药与旧药在不良反应、生命体征及IPI数据方面的差异, 建立了预测模型。

关键词

Pearson相关系数法, 随机森林, XGBoost, 集成模型, Mann-Whitney检验, SVM (支持向量机)

Analysis and Prediction of Therapeutic Effect of New Sedative Drugs Based on Combination Model

Junda Kang, Bo Wang, Yawen Xia, Youyan Wan*

School of Artificial Intelligence, Jiangnan University, Wuhan Hubei

Received: Dec. 12th, 2023; accepted: Jan. 7th, 2024; published: Jan. 15th, 2024

Abstract

Clinical research can verify scientific hypotheses and provide evidence-based medical evidence, which is not only an important guarantee for the development of clinical diagnosis and treatment methods and medical disciplines, but also a key link in the research of new drugs. This article studies the clinical trial data of an original sedative drug and a new sedative drug during surgery, analyzes the differences in adverse reactions, vital signs, and IPI data between the new drug and the old drug, and establishes a prediction model.

*通讯作者。

Keywords

Pearson Correlation Coefficient Method, Random Forest, XGBoost, Ensemble Model, Mann-Whitney Test, SVM (Support Vector Machine)

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 问题提出

本文问题提出基于 2023 年第十五届“华中杯”大学生数学建模挑战赛[1]中的 A 题，即为探究某药物研究中心的新型镇静药物临床实验疗效，附件中提供了 1245 名患者的 185 项指标数据，其中包括患者的基线特征、进药后的生命体征、术中术后的不良反应等数据。基于上述背景，分别解决以下三个问题。

问题 1：分析附件中关于术中、术后 24 h 不良反应，新药组和原有药物组是否存在显著差异；并根据患者基本信息和镇静药物种类，对患者术中、术后 24 h 的不良反应进行预判。

问题 2：分析新药组和原有药物组在生命体征数据方面是否表现出显著差异；若有显著性差异，分析出造成显著性差异的因素。

问题 3：根据患者的基本信息及用药信息，建立数学模型对其用药后三分钟内的 IPI 数据进行预测。

2. 模型假设

本文拟建立多种预测分析模型来解决以上三个问题。在建立模型之前，先提出模型成立的假设条件：假设患者所填写的问卷信息效度和信度分析均良好。

假设患者生命体征记录数据误差在合理范围内，不影响后续回归分析与预测。假设不平衡的问卷数据不会影响模型效果。假设空缺的数据不会对模型造成太大影响。

3. 符号说明

在本文中，使用了一些特定的符号来表示不同的概念和信息，具体符号意义如表 1 所示。

Table 1. Explanation of mathematical symbols

表 1. 数学符号说明

数学符号	说明
α	卡方检验显著水平
N_1	新药组
N_2	原有药组
n_1	N_1 组独立样本
n_2	N_2 组独立样本
U_α	Mann-Whitney 检验拒绝零假设的阈值
B_i	Lasso 回归中的特征变量名
b_i	Lasso 回归中特征变量的系数
p_i	患者术后满意值序列平均值

4. 模型的求解与结果分析

4.1. 问题 1 模型建立与求解

4.1.1. 数据预处理

(一) 术中情况分析

Step 1 术中不良反应特征值划分。

经观察,发现有 48 名患者在术中有除呛咳、体动以外的不良反应,且这些反应分为“打嗝”、“打呼”、“打嗝 + 打呼”三种,我们将具有“打嗝 + 打呼”不良反应的数据拆分为“打嗝”和“打呼”两条数据,因此,术中其他不良反应有“无”、“打嗝”、“打呼”三个特征值。

Step 2 呕吐次数异常值处理。

对于术后 24 h 出现呕吐反应但并未填写呕吐次数的患者,我们采用众数(1 次)填补缺失值的方法对他们的呕吐次数进行赋值;对于填写 451 号患者填写的“四五次”,我们将其呕吐次数赋值为 4 次;对于 986 号患者填写的“多次”,我们将其呕吐次数赋值为 5 次。

4.1.2. Pearson 卡方检验模型

卡方检验通过比较两项或多项频数,检测在一定显著性水平上实际频数与以某种理论模型或分布特征假设为基础的期望频数的差异度。对于术中、术后 24 h 不良反应与镇静药物种类之间的相关性:

$$\chi^2 = \sum \frac{(A - E)^2}{E} = \sum_{i=1}^k \frac{(A_i - np_i)^2}{np_i} \quad (1)$$

Step 1: 建立原假设 H_0 : 使用不同镇静药物对术中、术后 24 h 的不良反应没有影响,即镇静药物种类与不良反引发中无显著性差异。Step 2: 根据理论分布计算期望频数。Step 3: 根据实际频数和期望频数计算样本卡方值。Step 4 根据自由度和显著性水平 α 在卡方分布表中查找出对应卡方临界值。设置显著水平 α , 计算 P-value。若 P-value > α , 则接受原假设,反之,接受备择假设。

经过上述操作,取 $\alpha = 0.05$, 对术中不良反应的卡方检验结果如表 2 所示。

Table 2. Chi-square test results of intraoperative adverse reactions in the new drug group and the existing drug group
表 2. 新药组和原有药物组对术中不良反应的卡方检验结果

特征	例数	新药组数	原有药组数	χ^2	P	显著性
呛咳				2.97	0.095	
有	53	7	46			
无	1192	724	468			
体动				4.4	0.036	√
有	112	83	29			
无	1133	687	446			
术中其他				3.606	0.165	
无	1197	728	469			
打呼	10	8	2			
打嗝	39	35	4			

从上表可以看出：对于“体动”这一不良反应，其 P 值为 0.036 ($P\text{-value} < 0.05$)，水平上呈现显著性，拒绝原假设，即新药组和原有药物组存在显著性差异。而对于“呛咳”和“术中其他”这两种不良反应，其 P-value 均大于 0.05，接受原假设，即新药组和原有药物组不存在显著性差异。

(二) 术后 24 h 情况分析

表 3 为关于新药组和原有药物组对术后 24 h 不良反应的卡方检验结果。

Table 3. Chi-square test for postoperative adverse reactions between the new drug group and the existing drug group at 24 hours

表 3. 新药组和原有药物组对术后 24 h 不良反应的卡方检验

特征	χ^2	P	显著性
是否出现了恶心	35.475	0.000	√
是否出现了呕吐	3.542	0.060	
是否恶心呕吐都有	1.969	0.161	
是否出现了头晕	0.096	0.757	
是否出现了头昏	5.783	0.016	√
是否出现了头痛	0.112	0.738	
是否头晕头昏头痛都有	2.739	0.098	
是否出现了嗜睡	1.162	0.281	
是否出现了乏力	11.21	0.001	√
是否嗜睡乏力都有	2.698	0.100	
是否出现了腹胀	23.562	0.000	√
是否出现了腹痛	7.498	0.006	√
是否腹胀腹痛都有	0.056	0.813	
是否有其他不舒服的情况	0.207	0.649	

对于“恶心”、“头昏”、“乏力”、“腹胀”、“腹痛”这 5 个不良反应，其 P-value 均小于 0.05，水平上呈现显著性，拒绝原假设，即新药组和原有药物组存在显著性差异。而对于其余的不良反应，其 P-value 均大于 0.05，接受原假设，即新药组和原有药物组不存在显著性差异。

4.1.3. T 检验模型

独立样本 T 检验用于检验两分组数据是否存在显著性差异，由于出现呕吐次数为定量指标，所以我们采用独立样本 T 检验模型对其进行检验。建立原假设 H_1 ：镇静药物种类对呕吐次数没有影响，即镇静药物种类在呕吐次数上无显著性差异。

检验结果：新药组和原有药物组在出现呕吐次数上的均值分别为：0.066、0.152；显著性结果 P 值为 0.109 ($P\text{-value} > 0.05$)，因此统计结果不显著，接受原假设，说明新药组和原有药物组在出现呕吐次数上不存在显著差异。

第一小问结论：对于术中不良反应“体动”、术后 24 h 不良反应“恶心”、“头昏”、“乏力”、“腹胀”、“腹痛”，新药组和原有药组存在显著性差异。

4.1.4. 不良反应预判

Step 1: 数据预处理。

对于 198 号患者，其年龄值为空，故剔除该患者的临床实验数据。

Step 2: 特征编码。

根据第一小问的结论，筛选出术中、术后 24 小时有以上不良反应的患者的基本信息及其使用镇静药物的种类，为便于后续回归模型的建立，对非数值型特征值进行编码。

Step 3: 建立回归模型。

逻辑回归 Logistic 回归模型是对线性回归模型的结果做出的 sigmoid 映射，映射函数为：

$$g(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

该函数可将所有函数值映射到(0, 1)区间内，即将逻辑回归的结果转化为(0, 1)之间的概率值，由此构造分段函数，即可实现二分类问题。

随机森林[2]随机森林是将随机子空间与 Bagging 集成学习两种方法相结合的集成算法。随机森林由一系列相互之间不相关的的决策树组成森林，结果为多棵决策树投票表决的结果，故其分类结果比单棵决策树分类结果要理想很多。

XGBoost 模型 XGBoost 全称是 Extreme Gradient Boosting，又被称为极限梯度提升算法。XGBoost 模型集优秀的模型效果和迅速的运算速度于一体，在 ui 规模性上拥有超高的运算性能。XGBoost 模型是在传统梯度提升模型的基础上进行改进优化，通过多次迭代累，将多个弱评估器组合成一个强评估器，以获得更好的回归表现。其本身亦属于集成学习算法。

Step 4: 模型集成。

集成模型采用投票法的方式实现，使用 Scikit-learn 库中的 VotingClassifier 类，将逻辑回归模型、随机森林模型和 XGBoost 模型传入，我们将投票方式设置为“硬投票”，即直接取各模型的预测结果中得票最多的类别作为集成模型的预测结果。接着，我们使用 fit 方法在训练集上训练投票模型，并使用 predict 方法在测试集上进行预测。最后，我们计算集成模型预测准确率。

Table 4. Model accuracy

表 4. 模型准确率

不良反应 \ 模型类型	逻辑回归	随机森林	XGBoost 模型	集成模型
体动	0.9278	0.9178	0.9144	0.9305
恶心	0.9064	0.9011	0.8984	0.9118
头昏	0.9742	0.9628	0.9734	0.9806
乏力	0.9248	0.9197	0.9011	0.9278
腹胀	0.9492	0.9465	0.9412	0.9492
腹痛	0.9813	0.9786	0.9759	0.9813

通过表 4，可以看出逻辑回归模型、随机森林模型、XGBoost 模型在未经集成处理的情况下表现效果就已经很优秀了，各模型预测准确率均高于 89%。通过模型的集成，使得最终模型的预测准确率均高于 91%。

4.2. 问题 2 模型建立与求解

4.2.1. 数据预处理

附件中提供了收缩压(mmHg)、舒张压(mmHg)、PetCO₂ (mmHg)、呼吸(次/分)、SpO₂ (%)、心率(次/

分)及 IPI、MOAA/S 评分这八个生命体征指标,并记录了患者在诱导前后的一些时间段及手术后这些指标的具体数据值。经过观察,我们发现收缩压诱导后 20 min (sbp20)、MOAA/S 评分诱导后 20 min (moaas20)、舒张压诱导后 2.5 min (dbp025)的数据值缺失率均超过 85%,故我们将这些特征指标进行了删除。对于生命体征数据其他的缺失值,我们采用了均值填补的方式对其进行了处理。

4.2.2. 独立样本 Mann-Whitney 检验

独立样本 Mann-Whitney 检验(Mann-Whitney U 检验),也称为 Wilcoxon 秩和检验,是一种用于比较两个独立样本的非参数假设检验方法。这个检验方法的原理是将两组样本合并,按照观测值的大小顺序排列,然后计算每个样本中的秩次。然后,通过比较两个样本的秩次总和来确定它们是否来自于相同的总体分布。

Step 1: 建立原假设 H_2 : 在原有药组和新药组中第 P 个生命体征的特征集没有显著不同。Step 2: 设置拒绝零假设的阈值 $U_\alpha = 0.05$ 。Step 3: 对生命体征数据按照新药组 N_1 和原有药物组 N_2 进行独立样本 Mann-Whitney 检验。Mann-Whitney 检验统计量 U_1 、 U_2 的计算公式如下:

$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2} \quad (3)$$

$$U_2 = R_2 - \frac{n_2(n_2 + 1)}{2} \quad (4)$$

式中: U_1 、 U_2 的最小值用于与显著检验阈值 U_α 相比较,若 $U_{\min} < U_\alpha$, 则拒绝假设 H_2 , 即两特征之间存在差异。

经过上述步骤,我们筛选出 57 条拒绝零假设的生命体征项,即原有药组和新药组在这 57 个生命特征项下呈显著性差异。下表 5 为部分筛选结果。

Table 5. Partial vital sign items and their test results that reject the null hypothesis
表 5. 拒绝零假设的部分生命体征项及其检验结果

生命体征项	Mann-Whitney 检验统计量	Mann-Whitney 检验 P 值
petco200	197388.0	0.016193446380254637
spo200	170590.5	0.012084835724673337
petco2005	200366.0	0.003912119004831138
spo2005	175199.0	0.02415294628636724
moaas005	209493.5	2.682934692430315e-09
sbp1	148406.0	2.677851145756807e-08
dbp1	146318.0	3.589315945192707e-09
spo21	173621.5	0.003414797369753047
moaas1	129600.5	1.320753812046705e-27
sbpjnjing	151029.5	2.8199259043516447e-07

4.2.3. 造成在生命体征数据方面呈现显著性差异的因素

Lasso 回归是一种线性回归方法,它通过加入一个 L1 正则化项来对模型进行约束和选择。

Step 1: 分别以患者基本信息和镇静药物种类为自变量,具有显著性差异的生命体征项作为因变量建立 Lasso 回归模型。加入 L1 正则化后的 Lasso 损失函数式(5)所示

$$J = \frac{1}{n} \sum_{i=1}^n (f(x_i) - Y_i)^2 + \lambda \|\omega\|_1 \quad (5)$$

式中, n 为样本个数, λ 为惩罚值, $\lambda \|\omega\|_1$ 为 L1 正则化。

Step 2: 我们以训练集: 测试集 = 8:2 的比例对数据集进行划分。

Step 3: 确定 Lambda 值, 并基于最佳的 lambda 值建立 Lasso 回归模型(下以诱导前 1~3 min 的 PetCO₂ 为例)通过交叉验证, 得到不同 λ 取值下 Lasso 模型的可视化效果, 见图 1。

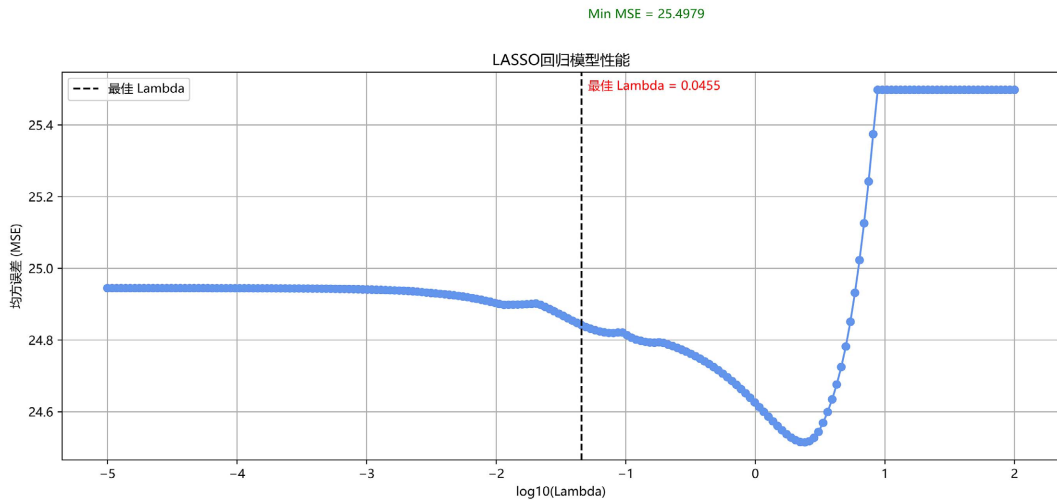


Figure 1. Cross validation rendering
图 1. 交叉验证效果图

对于图 1, 其横坐标代表 λ 的对数取值, 纵坐标代表拟合模型的均方误差。

图 2 为 Lasso 拟合系数轨迹, 其给出了在不同 λ 取值下各特征拟合系数值的变化情况及影响因素的筛选结果。

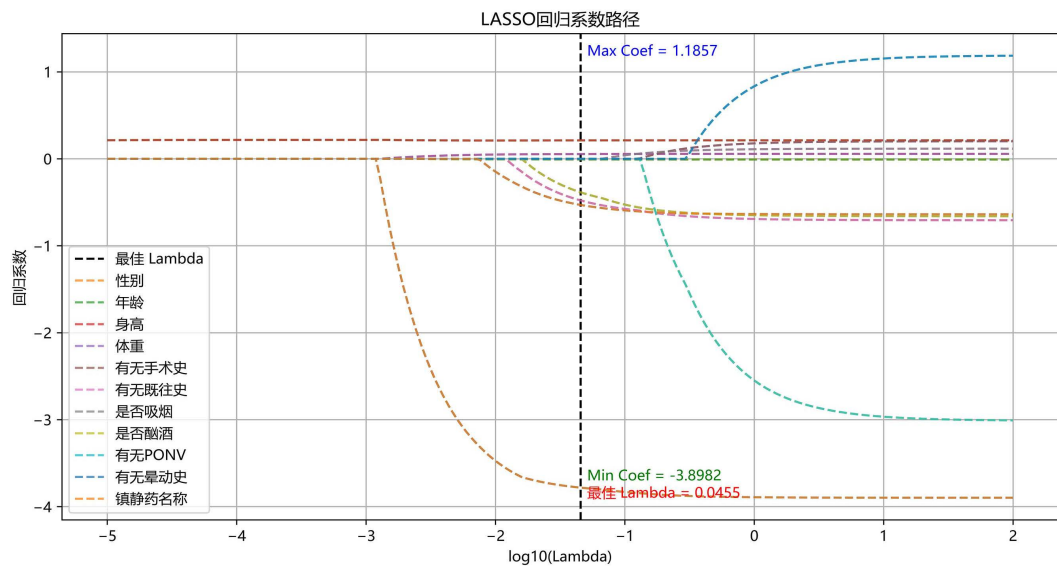


Figure 2. The relationship between different λ value and characteristic variable coefficient value
图 2. 不同 λ 取值与特征变量系数值之间的关系

为得到更精确的模型，我们选用误差最小时的 λ 取值，即 $\lambda = 0.0455$ ，此时的模型中包含特征变量数10个，将它们分别表示为 $B_i (i = 1, 2, \dots, 10)$ 并将其具体系数表示为 $b_i (i = 1, 2, \dots, 10)$ ，其中截距项 $b_0 = 48.536$ 。见表6。

Table 6. Characteristic variable coefficient value with minimum model error
表 6. 模型误差最小时特征变量系数值

自变量	系数值
截距	$b_0 = 48.5365$
镇静药名称	$b_1 = -0.6054$
有无手术史	$b_2 = 0$
有无 PONV	$b_3 = 0$
有无晕动史	$b_4 = 0$
是否酗酒	$b_5 = -0.0369$
有无既往史	$b_6 = -0.3365$
性别	$b_7 = -0.7037$
年龄	$b_8 = -0.0584$
身高	$b_9 = -0.0876$
体重	$b_{10} = 0.0844$

Step 4: 确定影响比重最大的因素。

Lasso 回归中的标准化系数(也称为 L1 范数)表示每个特征的重要性或权重大小。在 Lasso 回归中，标准化系数越高的特征表示其在模型中的贡献越大，因此可以认为它们与目标变量之间的关系更为密切，与之相关性更高。

我们找出每个模型标准化系数最大值对应的特征作为体征造成在生命体征数据方面呈现显著性差异的因素。仍以诱导前 1~3 min 的 PetCO₂ 为例，由表 5 可以得到各特征变量的系数之间的关系如式(6)所示：

$$b_7 > b_1 > b_6 > b_9 > b_{10} > b_8 > b_5 > b_2 = b_3 = b_4 \tag{6}$$

即对于诱导前 1~3 min 的 PetCO₂，选取“性别”作为造成在生命体征数据方面呈现显著性差异的因素。同上理，我们对第一小问中筛选出的 57 个生命体征项逐一进行 Lasso 回归分析，得到部分结果如下表 7 所示：

Table 7. Factor determination
表 7. 因素确定

生命体征	影响该值的最主要因素
诱导前 1~3 min PetCO ₂ (petco200)	性别
诱导前 1~3 min SpO ₂ (spo200)	性别
诱导后 0.5 min PetCO ₂ (petco2005)	性别
诱导后 0.5 min SpO ₂ (spo2005)	年龄
诱导后 0.5 min (moaas005)	镇静药名称

续表

诱导后 1 min 收缩压(sbp1)	镇静药名称
诱导后 7 min 心率(HR7)	有无晕动史
手术结束时心率(HRjieshu)	镇静药名称
手术结束时 MOAA/S 评分(moaasjieshu)	性别
手术结束时 PetCO ₂ (petco2jieshu)	镇静药名称

由此可见,造成在生命体征数据方面呈现显著性差异的因素不一定是新型镇静药物的使用,还有可能是性别、年龄、有无晕动史等。

4.3. 问题 3 模型建立与求解

4.3.1. 数据预处理

IPI 指标是刻画生命体征的核心指标。由前文的 IPI 指标分类可知, IPI 指标在区间(1, 10)内, 不同指标对应着患者不同的临床状态。根据题目要求, 我们将患者诱导后 0.5 min 的 IPI (IPI005)、诱导后 1 min 的 IPI (IPI1)、诱导后 1.5 min 的 IPI (IPI015)、诱导后 2 min 的 IPI (IPI2)、诱导后 2.5 min 的 IPI (IPI025)、诱导后 3 min 的 IPI (IPI3)这 6 项生命体征指标提取出来。对于缺失值, 我们采取众数填补; 对于 591 号患者, 其 IPI 值填写为 0, 数据异常, 故在本题将该患者的数据剔除。

4.3.2. 建立预测模型

本题欲采用下述三种数学模型对 6 项 IPI 指标进行预测:

SVM 模型 SVM 指支持向量机(Support Vector Machine), 其主要思想是将数据映射到高维空间中, 使得在该空间中能够找到一个最优的超平面, 将数据分成两个类别。在找到最优超平面的过程中, SVM 依赖于一些被称为“支持向量”的数据点, 这些点距离最优超平面最近。

决策树模型的基本思想是将数据分成多个小组, 每个小组具有相同的属性, 并且每个小组都属于一个类别。这种分组可以通过一系列的决策来完成, 每个决策都是基于数据的某个属性。决策树从根节点开始, 通过一系列分支和叶子节点来表示不同的决策路径。

在决策树算法中, 每个分支代表一个属性, 每个叶子节点代表一个类别。通过分裂数据, 决策树算法找到了最重要的属性, 这些属性能够帮助我们对数据进行分类。

BP 神经网络模型[3] BP 神经网络(Back Propagation Neural Network)是一种常用的人工神经网络模型, 由输入层、隐藏层和输出层组成。每个层都由一些神经元组成, 神经元之间通过加权连接相互作用。BP 神经网络使用反向传播算法来训练模型, 该算法可以根据误差信号将误差反向传播到每个神经元中, 然后使用梯度下降算法来更新每个神经元的权重和偏置。

BP 神经网络的训练过程包括两个阶段: 前向传播和反向传播。前向传播将输入数据传递到输出层, 计算输出层的误差, 然后将误差反向传播回隐藏层, 通过调整每个神经元的权重和偏置来更新模型。这个过程不断迭代, 直到模型收敛为止。

Step 1: 以用药信息和患者信息为自变量, 3 分钟以内的 IPI 数据作为因变量, 分别建立 SVM、决策树、BP 神经网络三种模型。

Step 2: 按照训练集: 测试集 = 8:2 来训练模型。

Step 3: 通过比较各个模型的评估参数, 找出对于各个阶段 IPI 数据的最优预测模型。

4.3.3. 最优模型的选择

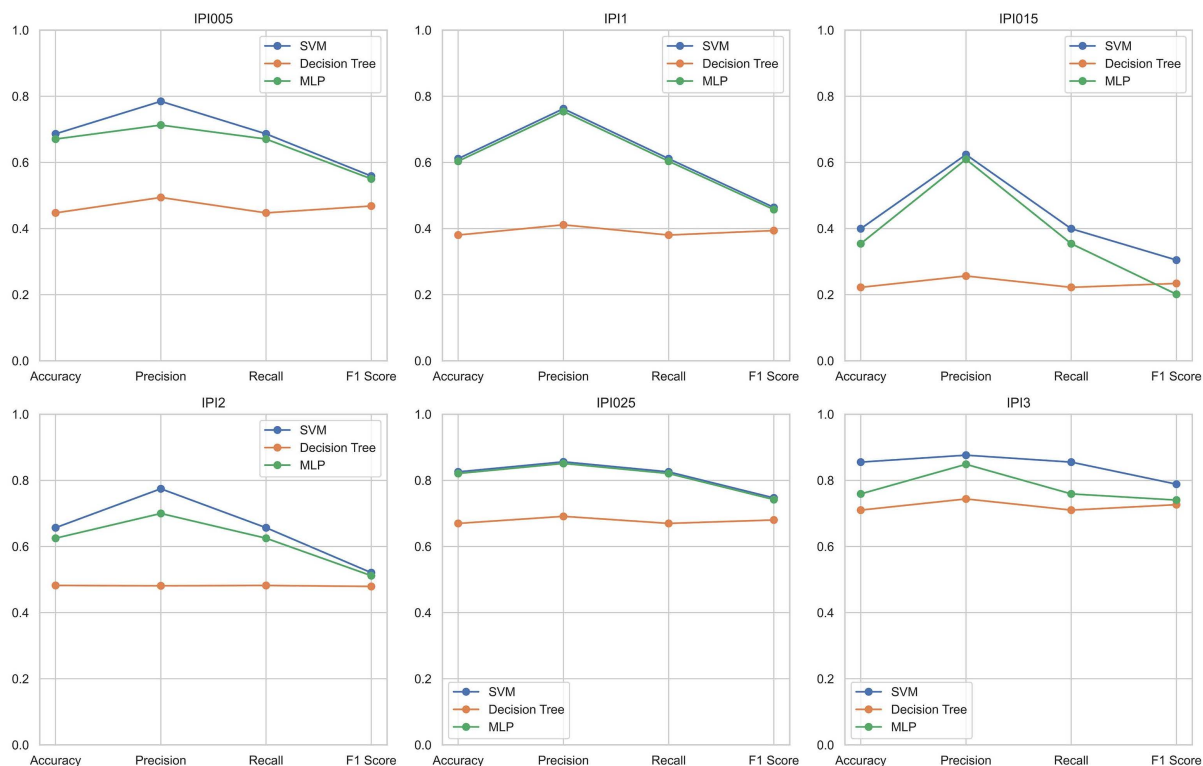


Figure 3. Evaluation indicators based on three prediction models

图 3. 基于三种预测模型的评价指标

其中 Accuracy、Precision、Recall、F1 均为评价模型的重要指标。Accuracy: 预测正确样本占总样本的比例, 准确率越大越好。Precision: 预测出来为正样本的结果中, 实际为正样本的比例, 精确率越大越好。Recall: 实际为正样本的结果中, 预测为正样本的比例, 召回率越大越好。F1: 精确率和召回率的调和平均, 精确率和召回率是互相影响的, 虽然两者都高是一种期望的理想情况, 然而实际中常常是精确率高、召回率低, 或者召回率低、但精确率高。若需要兼顾两者, 那么就可以用 F1 指标。

我们对所有模型均进行了以上四项评价指标的评估, 得到评估结果如图 3 所示。基于图 3 得到最优预测模型如表 8 所示。

Table 8. Determination of the optimal prediction model

表 8. 最优预测模型的确定

IPI 指标	预测模型	预测准确率
IPI005	SVM (支持向量机)模型	68.633%
IPI1	SVM (支持向量机)模型	61.126%
IPI015	SVM (支持向量机)模型	39.946%
IPI2	SVM (支持向量机)模型	65.684%
IPI025	BP 神经网络模型	82.574%
IPI3	SVM (支持向量机)模型	85.523%

对于指标 IPI025, BP 神经网络模型较 SVM 模型、决策树模型预测效果更好; 对于指标 IPI005、指标 IPI1、指标 IPI015、指标 IPI2 和指标 IPI3, SVM 模型的预测效果最好。

5. 模型的评价及推广

5.1. 模型评价

构建预测不良反应模型中通过集成模型的方法对不良反应进行预测, 融合多个模型的优点, 有效提高模型预测精度。使用数据的秩进行分析减少了异常值和离群值的影响并筛选出表现出显著差异的生命体征指标, 使回归分析的结果更加准确。在预测 IPI 数据中, 使用多模型和评价指标选出最优模型, 让预测更加精准。

构建集成模型中, 数据量需要足够大, 否则结果会失真, 此时通过合并一些类别, 以增加每个类别的样本数量, 从而更接近样本数量足够大的情况。当问题中存在大量相同值时, 再进行预测会导致信息丢失, 会影响筛选生命指标, 这种情况可以将相同的数值合并为一个秩次, 然后计算平均秩次。这样可以避免丢失太多信息。

5.2. 模型推广

对于此次两种镇静药物的临床实验疗效分析与预测有重要的现实意义。本文所提出的模型可以为临床医生在使用镇静药物时更好地预防和减轻不良反应的发生提供参考和建议, 使得临床医生更好地把握各种生命体征的变化情况, 及时调整镇静药物的用药剂量和种类, 以达到更好的临床治疗效果。

此外, 本研究的方法和模型也可以为相关学科领域提供一种有效的研究思路和方法, 例如药学、统计学和数据挖掘等领域, 可以参考我们所使用的数据预处理方法和建立的模型框架, 以及模型评价指标等方面的经验。

最后, 我们希望本文的研究成果可以为相关疾病的临床治疗和药物研发提供有益的参考和启示, 为人类健康事业做出一份微薄的贡献。

参考文献

- [1] 湖北省工业与应用数学学会. 第十五届华中杯大学生数学建模挑战赛赛题及数据[EB/OL]. <http://hzbmmc.com/views/award/award-item.html?navigate=inform&id=1652637360983003137>, 2023-04-30.
- [2] 任静莹, 马成满, 毕四旭, 邵喜高. 基于随机森林模型和遗传算法对抗乳腺癌药物的优化研究[J]. 鲁东大学学报(自然科学版), 2023, 39(2): 159-164.
- [3] 许美贤, 郑琰, 李炎举, 吴伟豪. 基于 PSO-BP 神经网络与 PSO-SVM 的抗乳腺癌药物性质预测[J]. 南京信息工程大学学报(自然科学版), 2023, 15(1): 51-65.