

# 研究人的血氧饱和度的评价预测模型

陈太龙 吕希瑞 何成杰

华北理工大学数学建模创新实验室 河北 唐山 063210

【摘要】本文研究了人体指标与血氧饱和度的关系，建立了表征一个人血氧饱和度水平和健康状况的评价模型和预测模型。首先为了建立血氧饱和度水平与健康状况的关系，采用因子分析法，分析了5项人体指标对血氧饱和度的影响，得到评价模型，对人体健康状况进行了评分。其次，利用灰色关联分析法得到与目标相关性较小的两个因子，然后建立正则化超限学习机模型进行两级预测。用关联度较小的指标预测剩余指标的对应值，再用剩余指标预测血氧饱和度的平均值。考虑到该预测模型可能存在误差，采用多元线性回归方程的预测进行对比验证，得到超限学习机的预测效果较好。

【关键字】因子分析；灰色关联度分析；超限学习机

## 1 建立健康评价模型

利用因子分析法<sup>[1]</sup>进行综合评价，根据不同因素得到每人的健康评分，将健康评分与血氧饱和度水平建立联系。通过查阅相关文献，确立影响人们的血氧饱和度水平的五个因素分别为年龄，BMI，性别，吸烟史，当前吸烟状况。整理36人的不同指标的测量数据，因子分析前，利用spss软件处理进行KMO检验，可得到五个指标均可进行因子分析。

因子分析中，求得各个公共因子的方差贡献率占总权重，再利用各个因子的权重汇总，可得评价得分的相关函数为：

$$Y_j = \alpha_1 F_{1j} + \alpha_2 F_{2j} + \dots + \alpha_k F_{kj} \quad (1)$$

其中， $\alpha_k$ 为第k个公共因子的方差贡献率占总权重， $F_{kj}$ 为第j个被评价指标在第k个公共因子的得分， $Y_j$ 为第j个被评价指标的综合得分。

我们规定一个人的健康的分值为0-10，将36个人从1-36进行编号，利用以上因子分析法，根据五个指标对它们进行健康评分，计算得到的健康值。

表1 每人的健康评分

个人序号	健康评分
1	9.6
2	8.5
3	9.4
...	...
36	8.9

## 2 预测模型的建立

### 2.1 灰色关联度分析

为预测模型确立变量。利用灰色关联度分析<sup>[2]</sup>求出不同因素影响人的血氧饱和度水平的程度，选取两个关联度最小的因素，为预测做准备。其关联度表达式为：

$$R_i = \frac{1}{n} \sum_{i=1}^n \xi_i(t) \quad (2)$$

其中 $R_i$ 为比较序列和参考序列的关联度，n为序列的长度。将数据代入模型运算，得到各个因素关联度结果如下图所示：

表2 各个指标的关联度

关联度	$r_1$	$r_2$	$r_3$	$r_4$	$r_5$
数值	0.4565	0.7894	0.4156	0.2316	0.7789

结合上表的结果进行分析，将关联度按从小到大的顺序进行排序，关联度越大，对目标的影响程度越大。选取两个关联度最小的因素分别为：性别，吸烟史。

### 2.2 基于正则化的超限学习机模型

为了定量分析不同指标对不同人的血氧饱和度水平的影响，利用基于正则化的超限学习机<sup>[3]</sup>来进行两次预测。一层是由关联度小的指标来预测其他指标对应的值，即关联度较大的因素。另一层是根据预测得到其他三个指标的值，来预测不同指标下的血氧饱和度的平均值。

极限学习机采用了单层前馈神经网络结构。输入样本总数为N，输入数据为X，网络输出为O，隐藏层输入权重矩阵为，隐藏层激活项为H，激活函数为Sigmoid函数，其中b为阈值。激活项H计算方程为：

$$H = g(\omega X + b) = \frac{1}{1 + e^{-(\omega X + b)}} \quad (3)$$

传统极限学习机存在不适定问题，即当输入矩阵X中的某些列之间存在很强的线性关系或者不是列满秩。在这种情况下，可以添加一个正则化项来解决，得到新的转换式：

$$\hat{\beta} = \left( H^T H + \frac{1}{C} I \right)^{-1} H^T T \quad (4)$$

其中，C为正则化系数，I为单位矩阵。通过正则化处理，最终得到正则化极限学习机模型，实现快速稳定的网络

输出。在第一层运算中得到系数矩阵分别为  $R_1, R_2$ 。

其中变量指标为  $A = (a_1, a_2)$ ，预测得到的血氧饱和度平均值  $Y_1$ ：

$$Y_1 = A \cdot R_1 \cdot R_2 \quad (5)$$

运用模型进行运算，最终得到的36人的血氧饱和度的实际数据，可预测数据以及每人的健康评分，结果如下表：

表3 模型预测结果

个人序号	预测的血氧饱和度	健康评分
1	98.45	9.6
2	92.46	8.5
3	96.23	9.4
...	...	...
36	93.66	8.9

由数据可得到血氧饱和度水平与健康程度的关系，当血氧饱和度越大时，健康评分越高。此预测模型可根据相关的参数的输入来表征一个人的血氧饱和度水平。

### 3 多元线性回归法

为了验证建立的预测模型的预测效果，用多元回归来对比。根据灰色关联度分析得到的主要因素，可作为变量来建立多元线性回归方程，来进行预测，由此一个人的血氧饱和度的值可表示为：

$$Y = \sum_{i=1}^5 y_i + \varepsilon \quad (6)$$

$\varepsilon$  为常数，需求解不同变量的系数， $y_i$  可表示为：

$$y_i = c_i a_i (i = 1, 2, 3, 4, 5) \quad (7)$$

其中  $k_i$  为方程的系数， $a_1, a_2, a_3, a_4, a_5$  为分别代表年龄，BMI，性别，吸烟史，当前吸烟状况这五个变量。由

SPSS 仿真结果得到方程为：

$$Y_2 = 0.5978a_1 + 0.4589a_2 + 1.4589a_3 + 2.3654a_4 + 3.6180a_5 + 64.2354 \quad (8)$$

将每人的指标代入回归方程中进行预测，得到的预测值结果：

表4 回归模型预测结果

序号	血氧饱和度预测值	健康评分
1	97.98	9.6
2	95.23	8.5
3	96.01	9.4
...	...	...
36	93.79	8.9

由此数据与超限学习机模型的预测结果对比可知，超限学习机模型预测效果更好，与实际值的拟合度更相近，建立的预测模型具有可行性。

### 4 结束语

将人的血氧饱和度与健康状况相联系，建立健康评价体系与预测模型，来研究了人的血氧饱和度的典型模式。其中利用灰色关联分析与正则化的超限学习机结合，进行改进。将指标分为两层来进行两层预测，得到更准确的结果。最后引入对比思想，验证所建立预测模型的效果更好，合理可行。

### 【参考文献】

- [1] 游家兴. 如何正确运用因子分析法进行综合评价 [J]. 统计教育, 2003(05):10-11.
- [2] 刘萍, 檀朝东, 刘畅. 基于灰色关联分析法的 SAGD 新井组注采参数的优选 [J]. 中国石油和化工, 2014, (12):53-55.
- [3] 张立佳, 赖晓平, 曹九稳. 正则化超限学习机的多分块松弛交替方向乘法 [J]. 模式识别与人工智能, 2019, 32(12):1107-1115.