

基于 Python 的服装人体第 7 颈椎点高度预测

尹 喆¹, 宁艳群², 何宇雯¹, 尚笑梅^{*1}

(1. 苏州大学 纺织与服装工程学院, 江苏 苏州 215006; 2. 湖南东方时装有限公司, 湖南 长沙 410000)

摘 要:人体生理结构中第 7 颈椎点骨骼不突出, 位置难寻找, 导致该部位高度测量困难。基于苏州大学人体测量实验团队获得的人体数据, 用 Python 语言对其进行数据预处理、数据挖掘以及回归分析。通过建立身高、侧颈点高、下颌高、腰高与第 7 颈椎点高的多元线性回归关系, 预测出第 7 颈椎点的高度, 并对回归模型进行评估和优化, 为当前非接触式人体测量仪在获取第 7 颈椎点高度时遇到的问题提供一种解决思路和研究方法。

关键词: 服装; 人体测量; 第 7 颈椎点; Python 数据处理; 多元线性回归

中图分类号: TS 941.17 **文献标志码:** A **文章编号:** 2096-1928(2020)03-0206-04

Height Prediction of the Seventh Cervical Vertebrae Point in Garments Based on Python

YIN Zhe¹, NING Yanqun², HE Yuwen¹, SHANG Xiaomei^{*1}

(1. College of Textile and Clothing Engineering, Soochow University, Suzhou 215006, China; 2. Hunan Dongfang Fashion Co., Ltd., Changsha 410000, China)

Abstract: Due to the physiological structure of the human body, the bone at the seventh cervical vertebrae point is not prominent, and the position is difficult to find, which makes it difficult to measure the height of this site. In this paper, the data preprocessing, data mining and regression analysis were performed in Python language based on the human body data obtained by the anthropometric experiment team of Soochow University. The height of the seventh cervical vertebrae was predicted by establishing a multiple linear regression among height, lateral neck height, mandibular height, waist height and seventh cervical vertebrae height, and the regression model was evaluated and optimized. It provides a solution and research method for the problems encountered by the current non-contact anthropometer in obtaining the height of the seventh cervical vertebrae.

Key words: clothing, anthropometry, seventh cervical vertebrae, Python data processing, multiple linear regression

服装人体测量学中,第 7 颈椎点较其他部位更为重要。根据 GB/T 16160—2008《服装用人体测量的部位与方法》^[1],由第 7 颈椎点展开的测量部位达 10 多个,例如,颈根围是用软尺经过第 7 颈椎点、侧颈点以及颈窝点量取颈根部一周的围长。在服装造型设计中,衣领是上衣的视觉焦点,连接着前后衣片,而其造型的创新设计离不开颈根围这一关键数据^[2-3]。

在采用传统手工测量的方法寻找第 7 颈椎点

时,通常让被测者低头,测量员用手去感知该部位凸出的骨骼点,并进行相关部位的测量。通过非接触式人体测量仪获取人体基准部位数据时,仪器水平投射光线存在扫描盲区,会造成人体点云模型中某些部位的缺失,虽然可以通过点云修补的方法让模型更完整,但在寻找部位关键点时仍然不能精确定位^[4-5]。在测量颈根围时,常用左右侧颈点、前窝点拟合的曲平面反求第 7 颈椎点,从围度进行求解^[6]。文中探索第 7 颈椎点与一些关键部位高度

收稿日期:2019-09-15; 修订日期:2020-02-20。

作者简介:尹 喆(1995—),男,硕士研究生。

* 通信作者:尚笑梅(1960—),女,副教授,硕士生导师。主要研究方向为服装数字化。Email:shangxiaomei@suda.edu.cn

值的关系,并建立回归模型,用已知部位高度预测第7颈椎点高度,为三维人体测量和建模提供参考。

1 研究方法

文中实验所用数据来源于苏州大学服装人体测量实验团队历年测体实验所获得的手工测量数据。

1.1 方法概述

Python 是一种面向对象的解释型高级编程语言,其结构简单,代码和语法定义清晰明确,易于学习和维护,可移植性和可扩展性非常强^[7-8];同时,Python 在数据分析和挖掘领域有许多现成的包和模块可以直接使用,能够极大地提高效率;除了语言内置库外,Python 还具有大量丰富的第3库,如 numpy, scipy, pandas, matplotlib, sklearn 等,包含了数据处理分析、统计分析、建模、图形可视化、数据挖掘与机器学习等几乎所有的数据分析功能^[9]。

数据挖掘是在庞大的数据库中通过各种特定算法来搜索和分析其中隐藏但潜在有用信息的过程^[10]。

回归预测是一种有条件的预测,依据估计出的回归方程,在给定自变量数值条件下,对因变量进行预测。依据相关关系中自变量的个数,回归预测划分为一元回归预测和多元回归预测。

1.2 测量对象

苏州大学 18~24 岁在校大学生,共计 210 人,每人测得 3 组数据。

1.3 样本选取

此次第7颈椎高预测实验选取了630个人体数据样本。原始数据共涉及10个人体测量部位,分别是身高、第7颈椎点高、左肩高、右肩高、侧颈点高、下颌高、腰高、臀高、内胯高、下肢高,目的是找出其中与第7颈椎点相关性高的部位,以预测第7颈椎点高度值,并建立它们之间的多元回归方程。

2 数据预处理与分析

2.1 格式处理

数据文件的读取采用 Python 第3方数据库 pandas 中的“.read_csv()”,故将文件格式转为.csv 文件,并将其中的数据标签属性全部改为英文(性别:sex,第7颈椎点高:jzg,身高:height,左肩高:zjg,右肩高:yjg,侧颈点高:cjdg,下颌高:xhg,腰高:yg,臀高:tg,内胯高:nkg,下肢高:xzg)。将性别属性更换成0-1型数据格式,男性置为“0”,女性置为“1”。

2.2 异常值识别与处理

通过 pandas 中“.describe()”来观测数据的整

体范围、大小以及波动趋势,发现不存在缺失值。但通过观察数据的描述性统计表格,发现某些部位最大值与最大值、最小值与最小值之间出现较大偏差。图1为每个部位的数据分布盒图,每个部位都有离群点存在。腰高、臀高、内胯高、下肢高4个部位离散程度较大,且出现腰高最小值小于臀高最小值、内胯高最大值高于臀高最大值的情况,不符合常理,推断是数据录入时存在错误,因此需将异常点剔除。

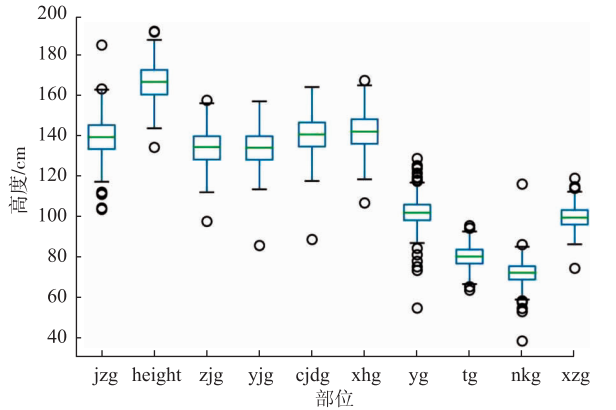


图1 原始数据分布盒图

Fig.1 Box of original data distribution

常用的异常值处理方法有 3σ 原则法和箱型图法。 3σ 原则指在数据服从正态分布的假设下,99%的数值应该位于均值3个标准差距离之内,超过此距离,就认为是异常值,该原则适用于数据充分大且离散程度较大的情况。箱型图能够反映数据的真实性,该方法判断异常值的标准是以四分位数和四分位距为基础,四分位数具有一定的鲁棒性,在多达25%的数据变得任意远时也不会很大地扰动四分位数,所以异常值不会对这个标准产生影响。箱型图识别异常值的结果比较客观,在识别异常值方面具有一定的优越性,因此文中实验采用箱型图法识别并剔除异常值,具体结果如图2所示。

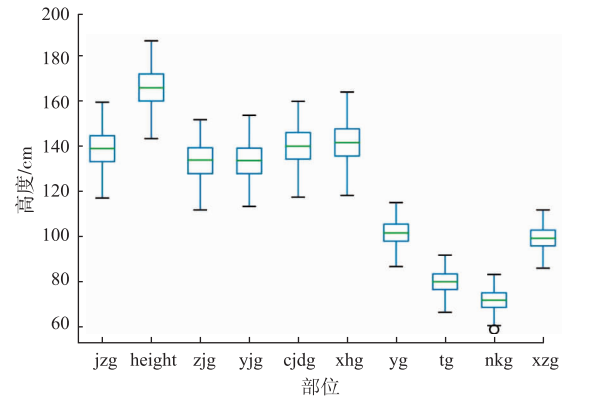


图2 预处理后的数据分布盒图

Fig.2 Box of pre-processed data distribution

由图 2 可以看出,经过数据异常值处理后,图 1 中每个部位的异常值都已经被剔除,数据分布正常,没有异常值点出现。预处理后的数据样本由原来的 630 个变为 598 个,其中男性数据 289 个,女性数据 309 个。

3 回归模型建立与优化

文中基于 Python 的第 3 方库 statsmodels 建立多元线性回归模型,statsmodels 库提供了许多能够在 Python 环境中自动调用的函数,如探索性数据分析类函数、统计类函数以及统计模型估计类和函数,这些函数可以用作线性回归分析、广义线性模型建立、统计检验、时间序列分析等。另外,statsmodels 库还具有强大的绘图能力,常用于展示数据分析的直观效果和诊断已构建模型的优良性。

3.1 多元线性回归模型建立

普通最小二乘法 (ordinary least squares, OLS) 是在线性回归模型中估计未知参数的线性最小二乘法。OLS 通过最小二乘法原则选择一组解释变量的线性函数参数,将给定数据集当中观察到的因变量(被预测变量的值)与预测变量之间残差的平方和最小化,在几何上,可解释为预测变量与因变量轴平行的平方距离之和,集合中每个数据点与回归表面上对应点之间的差异越小,则模型越适合数据^[11]。

从 statsmodels 库中引入 OLS,建立与第 7 颈椎点相关变量之间的多元线性回归模型^[12]。具体代码如下:

```
#建立回归模型
from statsmodels. formula. api import ols
formula = `jzg ~ height + zjg + yjg + cjdg + xhg +
yg + tg + nkg + xzg`
est1 = ols( formula, data = df2 ). fit()
print( est1. summary() )
```

根据各自变量对应回归系数的 t 检验 P 值 ($P > |t|$) 大小,可以判定各回归系数是否显著。表 1 为回归模型结果。表 1 中,如果给定显著水平 $\alpha = 0.05$,则左肩高、右肩高、臀高、内胯高、下肢高在显著水平 $\alpha = 0.05$ 下不显著。对于不显著的变量,予以剔除,并建立与剩余变量的线性回归模型。

建立第 7 颈椎点高与身高、侧颈点高、下颌高以及腰高的多元回归模型,核心代码如下:

```
#剔除不显著变量后
formula = `jzg ~ height - zjg - yjg + cjdg + xhg +
yg - tg - nkg - xzg`
```

```
est2 = ols( formula, data = df2 ). fit()
print( est2. summary() )
```

表 1 回归模型结果

Tab. 1 OLS regression results				
自变量	回归系数	标准差	t	显著性判定值
height	0.218 1	0.067	3.267	0.001
zjg	0.015 1	0.049	0.305	0.760
yjg	-0.051 8	0.064	-0.812	0.417
cjdg	0.201 3	0.044	4.569	0.000
xhg	0.316 5	0.057	5.566	0.000
yg	0.253 6	0.041	6.116	0.000
tg	0.054 8	0.064	0.855	0.393
nkg	0.013 6	0.061	0.221	0.825
xzg	0.041 9	0.051	0.826	0.409

表 2 为剩余变量回归模型结果。由表 2 可以看出,剔除不显著的部位后,余下的身高、侧颈点高、下颌高、腰高的显著性判定值均小于 0.05,都通过 F 检验和回归系数显著性检验,但该模型有较强的多重共线性问题。在模型建立之前,已将数据进行处理,并没有发现数据存在问题,原因可能是部位的选取过于接近且在同一个维度上,如身高、下颌高以及侧颈点高的位置过于接近,造成变量选取重复。但多重共线性不会影响拟合优度和预测优度,拟合值也不会受到影响,文中实验将不针对此问题进行深入讨论。

表 2 剩余变量回归模型结果

Tab. 2 OLS regression results of residual variables				
自变量	回归系数	标准差	t	显著性判定值
height	0.210 4	0.064	3.309	0.001
cjdg	0.201 1	0.039	5.207	0.000
xhg	0.326 0	0.056	5.867	0.000
yg	0.296 7	0.034	8.788	0.000

3.2 回归模型优化

上述 OLS 模型在对原始数据训练之前,未将数据集进行切分,这会导致预测模型仅适用于现有的数据集。一个模型如果只是重复它的样本标签,将获得一个完美的分数,但在陌生的数据集中,它无法预测任何有用的东西,这种情况称为过度拟合^[13]。为了避免过度拟合,在执行机器学习实验时,通常会将部分可用数据作为测试集 $x_{\text{test}}, y_{\text{test}}$,然后对测试集进行预测。

引入机器学习库 sklearn 中的评分函数“`.score()`”,对切分数据集前后的线性模型打分。核心代码如下:

```
from sklearn import linear_model
from sklearn. model_ selection import train_test_
```

```
split
x = df2[ ['height', 'cjdg', 'xhg', 'yg']]
y = df2[ 'jzg']
#未切分数据集
regr = linear_model.LinearRegression()
regr.fit(x, y)
y_pred = regr.predict(x)
print('Variance score: %.3f' % r2_score(y, y_
pred))
#切分数据集
x_train, x_test, y_train, y_test = train_test_
split(x, y, test_size = .4, random_state = 0)
regr = linear_model.LinearRegression()
fitted = regr.fit(x_train, y_train)
y_pred = regr.predict(x_test)
print('Variance score: %.3f' % r2_score(y_
test, y_pred))
```

上述代码运行结果显示,未切分数据集之前模型评分为 0.938,切分数据后评分为 0.947,比切分前高了约 0.01。图 3 为预测结果和真实结果比较。由图 3 可以看出,预测结果较好。实验选取切分数据集后的结果作为回归模型的参数,得到多元线性回归方程式:

$$y_{jzg} = 0.271x_{height} + 0.242x_{cjdg} + 0.180x_{xhg} + 0.373x_{yg} - 3.352$$

(1)

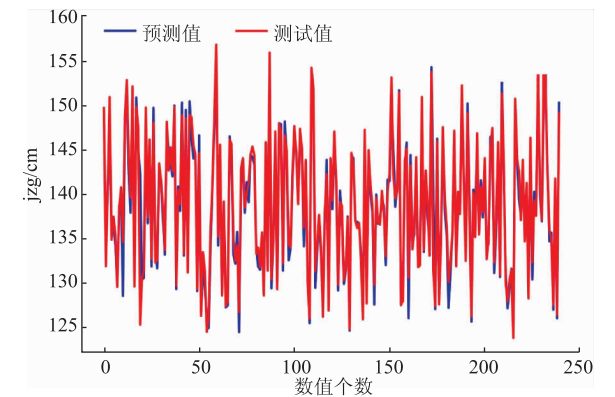


图 3 预测结果和真实结果比较

Fig.3 Comparison of predicted and real results

文中实验预测结果评分较高,且模型存在较强的共线性问题。可能原因如下:

- 1) 数据预处理过程中,对异常值过于严格,导致数据离散程度较小、集中度较高,但数据样本量少,预测处在不同身高档差范围内的人群时,模型的泛化能力不强,预测结果可能较差。
- 2) 人体各部位之间存在一定的相关性,用处在同一维度的不同部位来预测人体第 7 颈椎点高度,

可能出现影响因子冗余现象,而其他维度上的影响因子得不到检验,模型鲁棒性会降低。同时,通过显著性检验的部位间距较小,可能造成部位选取重复,模型出现拟合过度的情况。对于具有高度相关性、评分高的模型,模型优化效果不太明显。

4 结 语

服装用非接触式人体测量仪的测量精度和速度一直是技术人员追求和改进的目标。由于受人体复杂生理结构的影响,测量仪难以扫描到部分人体基准点,无法测得这些点的数据。文中以 Python 语言作为数据处理和分析的工具,利用 Python 中 statsmodels 库自带的 OLS 完成对人体第 7 颈椎点高度的预测,展现出其便利性与灵活性。实验得出身高、侧颈点高、下颌高、腰高这 4 个与第 7 颈椎点相关性较高的人体纵向尺寸,建立并优化了它们之间的多元回归模型,模型最终得分 0.947。该方法为非接触式人体测量仪在第 7 颈椎点高度的获取提供了一种较新颖的解决思路,有助于非接触式测量仪精度的提高。

参考文献:

[1] 全国服装标准化技术委员会. 服装用人体测量的部位与方法:GB/T 16160—2008[S]. 北京:中国标准出版社,2008.

[2] 王成泽,王建萍. 连翻领侧颈点造型影响因素探析[J]. 纺织科技进展,2011(5):75-78.
WANG Chengze, WANG Jianping. Study on influencing factors of side neck point in turn-down collar[J]. Progress in Textile Science Technology, 2011(5): 75-78. (in Chinese)

[3] 吴俊,蔡阳勇. 女装衣领结构设计方法研究[J]. 天津工业大学学报,2008,27(1):31-35.
WU Jun, CAI Yangyong. Pattern design of women's clothing collar[J]. Journal of Tianjin Polytechnic University, 2008, 27(1): 31-35. (in Chinese)

[4] 郭娟,羿莹. 非接触式三维人体扫描技术的应用分析[J]. 山东纺织科技,2018,59(1):39-41.
GUO Juan, YI Ying. Analysis on application of non-contact 3D human scanning technology [J]. Shandong Textile Science and Technology, 2018, 59(1): 39-41. (in Chinese)

[5] 周亮. 基于点云模型的人体尺寸测量方法[D]. 杭州:杭州电子科技大学,2015.

[6] 程达丽. 基于点云数据的人体特征尺寸提取[D]. 北京:北京服装学院,2012.

[7] 阮敬. Python 数据分析基础[M]. 北京:中国统计出版社,2017.

(下转第 269 页)