

运动手环的数据分析，以 fitbit 数据集为例

肖新元 *

江西机电职业技术学院 江西 南昌 330013

摘要：近年来，以智能手表、运动手环为代表的智能可穿戴设备发展乏力。文章利用开放数据集，使用当前流行的 Python 数据分析工具，对用户运动类型、用户活动、用户活动和睡眠、运动 Calories 回归等方面展开分析。利用可穿戴设备监测的数据分析结果表明，适度运动有利于身体健康，技术创新为智能可穿戴设备开拓光明的前景。

关键词：可穿戴设备；Python；机器学习；回归

随着工业化发展和电子产品普及，人们的生活方式相比十多年前有着非常大的变化。在我国，曾经随处可见的共享单车在帮助人们出行的同时，助力于人们运动和锻炼^[1]。如今骑行已经非常少见，而手机游戏成为危害广大青少年身体健康的主要因素^[2]。在上世纪 80 年代开始出现了智能的可穿戴设备^[3]，包括智能手表、运动手环、智能眼睛、智能衣物、头戴一体机等，这些智能设备集成了大量传感器，可以实时获取运动参数，包括运动步长、运动距离、睡眠状态、睡眠时长等；也可实时检测人体各种身体参数，包括心率、血压、血糖、血氧饱和度、睡眠状态等^[4]。借助于智能可穿戴设备，可为老人、中年、青少年等人群的健康状况提供数据支撑。

近年来，以智能手表为主的可穿戴设备销售额逐年提高，图 1 为 2015、2018、2021 年智能可穿戴设备市场销售额。IDC 提供的数据表明，2016 年全球智能可穿戴设备出货量增长 27.3%，2017 年的出货量增长 10.3%，至 2018 年全年可穿戴设备出货量为 1.722 亿台，同比增长 27.5%，2020 年全年整体出货量为 4.447 亿部，同比上升 28.4%；2021 年全年出货量为 5.336 亿部，比 2020 年增长 20.0%。2022 年全年出货量与 2021 年相比下降了 7.7%，这是该品类出货量的首年下降^[5]。闫龙雨等学者采用问卷调查方式对智能可穿戴设备未来发展趋势及应用对策进行了研究^[6]，王文韬等学者借助 SOR 理论对智能健康手环用户不持续使用行为进行研究^[7]。本文对智能可穿戴设备的运动相关数据进行充分地分析，帮助人们更好地了解自己的身体健康状况，促进人们保持健康的生活习惯，进行科学的运动。

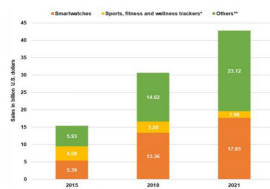


图 1 各类智能可穿戴设备的销售额

图 1 2015、2018、2021 年智能可穿戴设备市场销售额

1 分析基础

本文采用来自 kaggle 开放的数据集，该数据集在 2016 年 4 月 12 日-2016 年 5 月 12 日期间通过亚马逊 Mechanical Turk 进行的分布式调查的受访者获取。30 多名符合条件的 Fitbit 用户同意提交个人追踪器数据，包括身体活动、心率和睡眠监测等信息。

本文基于 Anaconda 进行数据分析和可视化，Anaconda 是一个开源的 Python 发行版，旨在简化 Python 环境的管理和部署。它包含了 Python 解释器以及大量常用的科学计算和数据分析库，如 NumPy、Pandas、Matplotlib、Seaborn 等。

2 运动分析

2.1 用户类型分析

这里，我们约定：平均每天运动总步长少于 5000 的称为静养运动类型用户，5000-7500 间称为轻度活动类型用户，7500-10000 间称为适度活动类型用户，10000-12500 间称为中度活动类型用户，超过 12500 的称为活跃活动用户。五种类型用户分别在属于“活跃活动”、“适度活动”、“轻度活动”、“静养活动”的活动时长如图 2 所示。静养活动类型用户的静养活动时长大，其他活动时长小，其他类型用户的活动时长基本与其所属类型一致。不难发现，中度活动类型的用户，因其在轻度、静养活动中的时间较长。

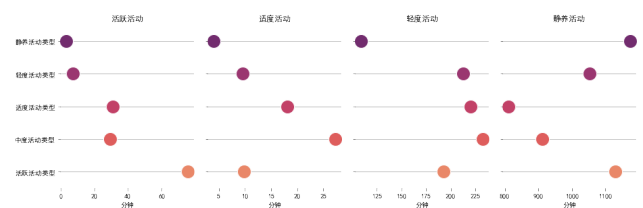


图 2 五种类型用户在不同情况下的活动时长

2.2 用户活动分析

人们的工作和生活一般受到时间周期性的影响，他们的活动表现出一定的规律性。数据记录了每个用户每天在“活跃活动”、“适度活动”、“轻度活动”、“静养活动”四种情形下的活动时长，图 3 展示了一周中每天各种类型活动的平均活

动时长。结果显示周六人们更愿意进行轻度活动，周日轻度活动则显著降低。

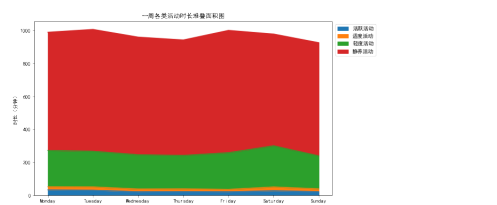


图3 一周各类活动时长堆叠面积图

数据集将用户的运动强度分为四个等级：0 久坐、1 轻度、2 适度、3 活跃。图4展示了一周每天分时的用户平均运动强度点图。可见周六8:00-14:00用户活动明显增强。

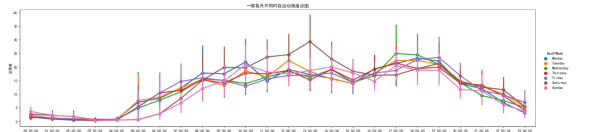


图4 一周每天不同时段运动总强度点图

2.3 运动睡眠分析

从医学角度来看，适量的运动可以帮助提高睡眠质量，并对整体健康产生积极影响。适量的运动可以促进身体的新陈代谢，增加体温，消耗能量，以及释放身体内的压力和紧张感。这些因素有助于改善睡眠，使人更容易入睡并获得更深沉的睡眠。此外，适量的运动还有助于调节生物钟，提高睡眠的规律性。过度劳累和过度激活的神经系统可能导致难以入睡和睡眠质量下降，因此过度运动可能对睡眠产生负面影响。但每天静养运动时间越长，不利于睡眠。图5表明静养运动与睡眠时长呈现负相关，图6表明，适度地运动不会影响睡眠质量。

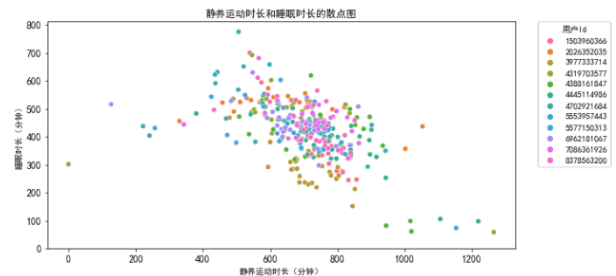


图5 用户静养运动时长与睡眠时长的散点图

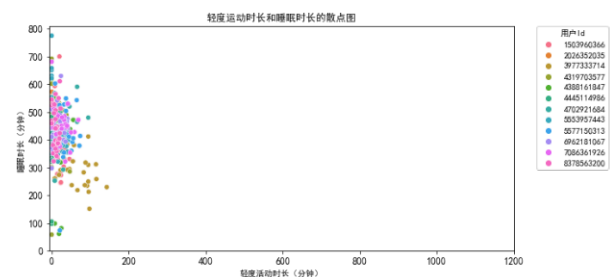


图6 用户轻度运动时长与睡眠时长的散点图

2.4 回归分析

人体卡路里是指单位能量的计量单位，用于衡量食物摄入

和能量消耗。通过摄入食物获取能量，而人体通过运动和代谢消耗的能量。通过计算摄入的食物中的卡路里含量，可以帮助人们控制自己的能量摄入，通过基础代谢和运动消耗卡路里，从而管理体重和健康，维持健康的身体素质和生命活力。

人体运动燃烧卡路里量与各种因素有关，为分析影响消耗卡路里各种因素，接下来搭建一个模型，对每天卡路里的消耗进行回归分析。将数据列Calories作为目标，去除Id和ActivityDate列，增加“星期”列，这里采用sklearn的LabelEncoder对“星期”列进行编码。采用机器学习的经典回归算法：线性回归、决策树回归、岭回归、贝叶斯岭回归、最近邻回归等进行分析，使用sklearn的均方差来计算测试集的RMSE，最后得到模型在不同回归算法下的均方根误差如表1所示。结果表明，对于运动数据的模型进行回归分析，采用线性回归和岭回归算法性能最优。

表1 不同回归算法的RMSE值

回归算法	线性回归	决策树回归	岭回归	贝叶斯岭回归	最近邻回归
RMSE值	359.952536	373.128953	359.888769	531.439953	604.862053

3 结束语

当前，智能可穿戴设备的出货量增长乏力，发展前景有待观察。但工作和生活压力促使人们关注身体健康，对相关设备需求依然旺盛。通过对Fitbit跟踪器数据进行分析，我们发现，用户活动类型分类、用户活动分析、睡眠分析、运动健康回归分析可帮助人们理解运动与健康关系等方面具有显著的说服力，因而对其他可穿戴设备产品的营销活动有一定的支撑作用。同时，设备生产厂家还需从产品的佩戴舒适性、功能扩展性、相关配套软件易用性等方面加大研发投入，比如情绪监测、AI助手等，相信智能可穿戴设备的发展前景依然光明。

参考文献

[1] 夏晨阳, 杨靖瑶. “共享单车”的普及对全民健身运动的影响研究[J]. 当代体育科技, 2019(11): 185-186.

[2] 陆舟. 手机游戏对当代儿童的危害分析[J]. 江西教育: 综合版(C), 2018(9): 15-16.

[3] 江洪, 张晓丹. 国内外智能可穿戴设备行业状况浅析[J]. 新材料产业, 2016(2): 2-7.

[4] SwaT+. 智能健身之选: FitbitInspireHR 智能手环[J]. 微型计算机, 2019(20): 28-29.

[5] 立鼎产业研究网, 2023-03.

[6] 闫龙雨, 汤洁. 智能穿戴设备未来发展趋势及应用对策研究[J]. 西部皮革, 2021(22): 125-126.

[7] 王文韬, 张震. 融合SOR理论的智能健康手环用户不持续使用行为研究[J]. 图书馆论坛, 2020(5): 92-102.