

第九届湖南省研究生数学建模竞赛承诺书

我们仔细阅读了湖南省高校研究生数学建模竞赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们完全清楚，在竞赛中必须合法合规地使用文献资料和软件工具，不能有任何侵犯知识产权的行为。否则我们将失去评奖资格，并可能受到严肃处理。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们授权湖南省研究生数学建模竞赛组委会，可将我们的论文以任何形式进行公开展示（包括进行网上公示，在书籍、期刊和其他媒体进行正式或非正式发表等）。

我们参赛选择的题号是（从组委会提供的赛题中选择一项填写）：

我们的参赛编号（请填写完整参赛编号）：202418001017

所属学校（请填写完整的全名）：国防科技大学

参赛队员（打印后签名）：1. 薛玉洁

2. 邓永林

3. 杨天池

指导教师或指导教师组负责人（打印后签名）：

日期：2024年7月12日

（请勿改动此页内容和格式。以上内容请仔细核对，如填写错误，论文可能被取消评奖资格。）

第九届湖南省研究生数学建模竞赛

题 目：基于智能手机传感器数据的人体活动状态识别模型

摘要

随着智能手机的普及，利用智能手机内置传感器记录和评估用户日常活动状态已成为现实。智能手机上的运动健康软件可以根据用户的日常活动，如跑步、步行、骑车、爬楼等，计算其每日消耗的热量。这些计算依赖于智能手机中的加速度计和陀螺仪，这两个传感器分别用于测量手机的线性加速度和角速度变化。

本研究旨在利用智能手机传感器数据建立模型，以准确识别用户的活动状态。对于这 10 余名实验人员运动状态所产生的数据，我们采用了多种数据预处理技术和机器学习算法，对这些活动数据进行了分析和建模。通过对模型的训练和测试，我们验证了模型在识别不同活动状态方面的准确性和鲁棒性。本文的研究结果显示，智能手机传感器数据在活动识别领域具有较高的应用潜力，能够为用户提供准确的活动状态监测和热量消耗计算。

针对问题一：对于 3 名实验人员活动状态数据进行分析，活动状态的关键在于准确识别不同的运动模式。考虑到加速度计和陀螺仪数据的变化、不同活动状态的数据特征、以及分类算法的准确性，本文建立了数据特征提取、特征选择等指标，将活动状态分类的问题转化为无监督聚类问题，随后基于机器学习思想，建立了 k -means 聚类模型和均值哈希模型。该模型充分考虑了加速度计和陀螺仪数据的变化、不同活动状态的特征影响，本文使用 Python 中的 scikit-learn 库对该模型进行求解，求解结果见表 4。

针对问题二：该问的关键在于从附件 2 提供的 10 名实验人员的活动数据中提取典型特征，并建立有效的判别模型。考虑到加速度计和陀螺仪数据的多样性和复杂性，本文建立了多种指标，将活动状态判别的问题转化为多分类问题。随后基于机器学习思想，建立了 XGBOOST 分类模型，并导出了判别树。该模型充分考虑了加速度、角速度和时间序列特征的影响。本文使用 XGBOOST 分类算法对该模型进行求解，并使用 scikit-learn 框架进行实现。求解结果见表 5。

针对问题三：对于实验人员活动状态数据进行分析，关键在于识别不同人员在相同活动状态下的数据差异。考虑到实验人员的年龄、身高、体重等因素的影响，本文建立了描述活动状态和生理特征的指标，将实验人员活动状态识别的问题转化为多因素数据分析的问题，同时本文结合了前两问的 k -means 聚类 and 均值哈希方法。随后，基于相关性分析思想，建立了多元回归。该模型充分考虑了年龄、身高、体重等因素的影响，本文使用多元线性回归算法（基于 scikit-learn 框架）对该模型进行求解，求解结果见表 6。

最后，我们对提出的模型进行全面的评价：本文的模型贴合实际，能合通过不同人员的活动状态来实现活动种类的预测，并且不同人员在相同活动状态下的传感器数据确实存在差异，这些差异与年龄、身高、体重等生理特征有一定关系。通过对传感器数据进行深入分析，可以实现基于活动状态数据的人员画像，为个性化健康管理和运动分析提供参考。

关键词：人体活动状态 k -means 聚类 均值哈希 多分类模型

目录

题 目：基于智能手机传感器数据的人体活动状态识别模型.....	I
摘要.....	I
1 问题综述.....	1
1.1 问题背景.....	1
1.2 问题提出.....	1
1.3 资料条件.....	2
2 模型假设与符号说明.....	3
2.1 模型基本假设.....	3
2.2 符号说明.....	3
3 数据预处理.....	4
3.1 附件 1 数据处理.....	4
3.1.1 指标选取.....	4
3.1.2 数据清洗.....	4
3.1.3 数据规约.....	4
3.2 附件 2 数据处理.....	5
3.2.1 指标选取.....	5
3.2.2 数据清洗.....	5
3.2.3 数据规约.....	5
3.3 附件 3 数据处理.....	5
3.3.1 指标选取.....	5
3.3.2 数据清洗.....	5
3.3.3 数据规约.....	5
3.4 附件 4 数据处理.....	6
3.4.1 指标选取.....	6
3.4.2 数据清洗.....	6
3.4.3 数据规约.....	6
3.5 附件 5 数据处理.....	6
3.5.1 指标选取.....	6
3.5.2 数据清洗.....	6
3.5.3 数据规约.....	6
4 问题一的分析与模型建立.....	6
4.1 问题分析.....	6
4.2 模型建立.....	7
4.3 模型求解.....	11
5 问题二的分析与模型建立.....	12
5.1 问题分析.....	12

5.2 模型建立	13
5.3 模型求解	13
6 问题三的分析与模型建立	14
6.1 问题分析	14
6.2 模型建立	15
6.3 模型求解	16
7 模型应用及分析	16
7.1 k -means 聚类策略	16
7.2 均值哈希策略	16
7.3 k -means 效率分析	17
7.4 均值哈希稳定性分析	17
8 模型评价与推广	17
8.1 模型的优点	17
8.2 模型的不足	17
8.3 模型的推广	18
参考文献	19
附 录	20
附录 A: 主要程序/关键代码	20

1 问题综述

1.1 问题背景

随着智能手机的普及，智能手机面临着诸多挑战，如：同类产品数量众多、种类丰富、价格低廉等。有资料显示，这类运动健康软件对手机使用人消耗的热量的计算依赖于其携带的智能手机记录其每天活动状态的数据。在智能手机需求不断扩大，而计算能力和数据资源又十分有限的背景下，传统的测量方式容易产生数据误差、精度不足等问题，严重影响了用户体验。因此我们需要寻找新的方式给出合理的活动状态评估方案。

目前，已经有很多学者对智能手机测量人体活动状态面临的问题进行了研究，也提出了很多算法来解决此类问题。传统的算法有加速度算法、陀螺仪算法等，近几年由于高性能计算设备的快速发展，还出现了深度学习算法、神经网络算法等。虽然这些算法可以快速得到活动状态的评估，但由于采用的是通用性框架，只能保证大多数情况下的合理性，个别情况可能出现误差较大的问题，这也促使我们运用更好的方法建立更加合理的模型。

智能手机测量人体的活动状态主要依靠内置于其中的运动传感器——加速度计和陀螺仪来实现。加速度计是用于测量手机在三个轴向（X，Y，Z）上的线性加速度变化情况的传感器，当手机发生加速度变化时，加速度计会感应到这种变化并将加速度数据传输给手机处理器进行存储或分析。陀螺仪是用来测量手机绕着三个轴向（X，Y，Z）旋转的角速度的传感器，当携带手机的人发生转向时，陀螺仪会感知到转向的速度和方向，并将这些数据传输给手机处理器进行存储或处理。通过加速度计和陀螺仪的数据，智能手机处理器可以感知到手机的姿态、角度和方向的变化，并设计相应的算法对手机使用人的活动模式进行判断或跟踪。

1.2 问题提出

为了验证上述方法的有效性，附件中邀请了 10 余名实验人员携带运动状态传感器进行活动，并收集他们的日常活动状态的数据。规定他们需要完成“向前走、向左走、向右走、步行上楼、步行下楼、向前跑、跳跃、坐下、站立、躺下、乘坐电梯向上移动、乘坐电梯向下移动”12 种活动，每种活动记录了 5 组实验数据，每组数据记录其数秒的线加速度和角加速度数据。

本文将尝试利用这些数据，建立模型完成活动状态识别和能量消耗估算的问题，具体包括以下三个问题：

- (1) 问题 1：附件 1 中有 3 名实验人员的运动数据，包含每名实验人员每种活动状态的 5 组加速度计和陀螺仪数据，但实验时未记录数据所代表的活动状态，请根据附件 1 提供的活动数据（每人 60 组数据），对每一位实验人员的活动状态的数据进行分类，在论文中将分类结果（编号）填入表 4。
- (2) 问题 2：附件 2 中有 10 名实验人员的活动数据，包含每位实验人员每种活动状态的 5 组加速度计和陀螺仪数据。实验时记录了每组数据所代表的实验人员的活动状态。请根据附件 2 提供的活动数据（每人 60 组数据）提取 12 类人员活动状态的典型特征，建立人员活动状态的判别模型，并利用该模型开展以下验证工作：①进一步运用问题 1 的分类模型对这 10 名实验人员数据进行分类（此时，不考虑实验人员的活动状态标签），比较问题 2 中的判别模型和问题 1 的分类模型的结果，分析采用分类模型对不同活动类型分类时的分类准确度。②附件 3 中收集有某实

验人员 30 次活动的状态数据，请运用你们的判别模型，给出该人员的活动状态，并在论文中将结果填入表 5。

- (3) 问题 3：附件 4 提供了参与问题 1 和问题 2 实验的 13 位实验人员的年龄、身高、体重等数据。请分析不同人员在相同活动状态下是否存在差异，并探讨活动状态数据与实验人员的年龄、身高、体重之间的关系，判断是否可以使用活动传感器数据进行人员画像。进一步，附件 5 中提供了问题 2 的 10 位实验人员中的 5 位某次活动数据，数据包含了每位实验人员在 12 类活动状态下的记录。请使用你们的模型，判断这些数据分别最可能来源于问题 2 中的哪一名实验人员，并在论文中将判别结果填入表 6。

1.3 资料条件

本题的三个问题均有附件，附件 1~5 中邀请了 10 余名实验人员携带运动状态传感器进行活动，并收集他们的日常活动状态的数据。规定他们需要完成“向前走、向左走、向右走、步行上楼、步行下楼、向前跑、跳跃、坐下、站立、躺下、乘坐电梯向上移动、乘坐电梯向下移动”12 种活动，每种活动记录了 5 组实验数据，每组数据记录其数秒的线加速度和角加速度数据。

对于附件 1，对应于问题 1 的求解，要求对三个人的运动状态数据进行 12 中类别的分类；对于附件 2~3，这是问题 2 中建立和验证人员活动状态判别模型的基础；对于附件 4~5，提供了 13 位实验人员的不同人员画像，要求对附件 5 中的活动状态归属者进行判断。以下对五个附件的内容及其格式进行详细解读。

(1) 附件 1：附件 1 中包括三个文件夹，每个文件夹对应 1 个实验人员。每个文件夹中包含了 60 组数据，每组数据中包含了 2000~5000 个人员活动状态数据。实验人员的活动状态数据存储在一个 Excel 文件中，包含以下 6 列：acc_x、acc_y、acc_z、gyro_x、gyro_y、gyro_z。这些列分别表示加速度计和陀螺仪在 X、Y、Z 三个轴向的记录值。在实验过程中，运动传感器被固定在实验人员的右下腹部，设备的采样率为 100Hz。加速度计的测量范围为 $\pm 6g$ ，陀螺仪的测量范围为 ± 500 度每秒（degree per second）。加速度计和陀螺仪的三个轴向中，X 轴沿重力方向，Y 轴沿实验人员的前进方向，Z 轴垂直于 X 平面并指向身体一侧。

(2) 附件 2 中的文件命名规则为“aitj.xlsx”，其中 i 表示活动类型，j 表示实验次数。例如，“ait2.xlsx”表示测量该实验人员第 1 类活动状态的第 2 组实验数据。活动状态编号如表 1 所示。

表 1 活动状态编码

活动状态	编号	活动状态	编号
向前走	1	跳跃	7
向左走	2	坐下	8
向右走	3	站立	9
步行上楼	4	躺下	10
步行下楼	5	乘坐电梯向上移动	11
向前跑	6	乘坐电梯向下移动	12

- (3) 附件 3：附件 3 提供了 30 组数据，每组数据的格式同附件 1；

(4) 附件 4 中提供了实验人员编号、年龄、体重、身高等四个信息；

(5) 附件 5：附件 5 中提供了五个未知人员的活动状态，其中每个人员的活动状态数据格式同附件 2，每组数据同附件 1；

2 模型假设与符号说明

2.1 模型基本假设

- (1) 为了简化模型分析，确保各个实验人员的活动数据可以单独处理，本文假设每位实验人员在每种活动时，其数据是独立的，即不同实验人员之间的数据没有相互影响，满足独立性假设；
- (2) 假设每位实验人员在相同活动时，其传感器记录的数据具有一致性和可比性，即使同一活动在不同实验人员之间存在差异，这些差异是可度量和可分析的，满足数据一致性假设；
- (3) 假设智能手机内置的加速度计和陀螺仪具有足够的精度，能够准确测量人体的活动状态变化、真实反映实验人员的运动情况；
- (4) 假设每位实验人员在同一活动状态下，其加速度和角速度特征在多次实验中保持相对稳定；
- (5) 对于问题 3，假设实验人员的年龄、身高、体重等个体特征与其活动状态数据之间存在某种线性或可线性化的关系。通过分析这些关系，可以有效地使用传感器数据进行人员画像和活动状态识别。

2.2 符号说明

本文定义了如下 15 个使用次数较多的符号，其余符号在使用时注明。

表 2 符号说明

符号	含义	单位
acc_x	加速度计在 X 轴向的记录值	g
acc_y	加速度计在 Y 轴向的记录值	g
acc_z	加速度计在 Z 轴向的记录值	g
$gyro_x$	陀螺仪在 X 轴向的记录值	dps
$gyro_y$	陀螺仪在 Y 轴向的记录值	dps
$gyro_z$	陀螺仪在 Z 轴向的记录值	dps
t	时域中的一维数据	
t_i	时域中单个的数据	
f	频域中的一维数据	
f_i	频域中单个的数据	
N	一维数据的长度	
x, y, z	X、Y、Z 轴	
mag	幅值	
$Median$	中位数	
E	期望	

3 数据预处理

3.1 附件 1 数据处理

3.1.1 指标选取

对于附件 1 中的数据，已经给定了以下主要指标：加速度计数据: acc_x 、 acc_y 、 acc_z ，陀螺仪数据: $gyro_x$ 、 $gyro_y$ 、 $gyro_z$ 。这些指标分别记录了实验人员在 X、Y、Z 三个轴向上的加速度和角速度变化，能够全面反映实验人员的活动状态。

进一步地，由于每组数据都是时间序列数据，并且存在以上六个维度的特征，根据文献[1]，我们分别选取了时域和频域两个维度的 25 个特征。其中的特征主要有均值、标准差、绝对差中位数、最大值、最小值、正数占比、信号幅度面积（轴向）、信号幅度面积（幅值）、能量、四分位距等。由于指标一共有 6 个，所以每组数据的所有特征一共有 150 个。

3.1.2 数据清洗

附件 1 中的数据较多，因此需要通过规范的数据清洗对数据进行整理。本文的数据清洗过程主要包括以下步骤：

（1）缺失值处理：检查数据集中是否存在缺失值。如果发现缺失值，可以采取删除含有缺失值的行，或者用均值、中位数等方法进行填补。

（2）异常值检测与处理：对每列数据进行异常值检测，判断是否存在明显偏离正常范围的值。例如，加速度计的测量范围为 $\pm 6g$ ，陀螺仪的测量范围为 ± 500 度每秒，超过这个范围的值可以认为是异常值。使用统计方法（如 3σ 原则）或箱线图法来检测异常值，并对异常值进行处理（如删除或替换）。

（3）噪声过滤：使用滤波器（如低通滤波器）对传感器数据进行噪声过滤，以平滑数据，减少测量误差的影响。

3.1.3 数据规约

由于本题中的数据单位均不同，因此需要进行数据规约操作，其包括以下步骤：

（1）数据分段：由于每组数据包含 2000~5000 个记录点，需要将数据进行分段处理。可以将每组数据分为若干固定长度的时间窗口（如 1 秒钟的窗口，包含 100 个数据点）。

（2）特征提取：从每个时间窗口中提取特征，如以下几种常用特征：

时间域特征有均值（Mean）、标准差（Standard Deviation）、最大值（Max）、最小值（Min）、四分位距（Interquartile Range, IQR）；

频率域特征有快速傅里叶变换（Fast Fourier Transform, FFT）、主要频率分量（Main Frequency Component）；

（3）数据归一化：对每个特征进行归一化处理，使其值在一个标准范围内（如 0 到 1 或 -1 到 1），以消除不同量纲之间的影响。

通过上述步骤对附件 1 中的数据进行指标选取、数据清洗和数据规约，可以获得高质量的特征数据，为后续模型训练和分析提供坚实的基础。

3.2 附件 2 数据处理

3.2.1 指标选取

与附件 1 一样，附件 2 中的每个文件包含了线加速度和角加速度的数据，我们沿用附件 1 中对原始数据的指标选取。

3.2.2 数据清洗

首先，对缺失值进行处理，检查了数据中的缺失值，使用插值或者删除缺失部分的数据。

之后，对异常值进行处理，检测和异常值，这可能是由设备故障或操作失误引起的异常数据点。可以采用统计方法或领域知识进行识别和修正。

3.2.3 数据规约

针对附件 2，我们进行了如下的数据规约：

（1）特征选择：根据活动的性质和分析的需求，选择最具代表性和区分性的特征进行后续分析。可能会采用特征工程的方法，如时域特征、频域特征或统计特征。

（2）数据转换：对数据进行必要的转换，如时间序列数据可能需要进行平滑处理或者时序分析。

（3）数据标准化：将数据进行标准化处理，确保不同数据之间的量纲一致，以便于后续的建模和分析。

在整个过程中，需要结合数据分析的具体目标和数据质量的实际情况，采用合适的方法和工具来保证数据的准确性和可靠性，从而为后续问题的求解提供可靠的基础。

3.3 附件 3 数据处理

3.3.1 指标选取

从每组数据中提取加速度计和陀螺仪的 X、Y、Z 轴向的记录值（acc_x、acc_y、acc_z、gyro_x、gyro_y、gyro_z）作为主要分析指标。这些数据能够反映实验人员的运动状态，并作为分类和建模的基础。

3.3.2 数据清洗

首先检查数据中的缺失值和异常值，并对其进行处理。缺失值可以使用插值法填补，异常值则需要通过统计方法或领域知识进行识别和修正。此外，确保所有数据的单位和范围一致，剔除不符合采样率或测量范围的数据点。

3.3.3 数据规约

为了减少数据的维度和复杂度，可以对原始数据进行特征提取，如计算每组数据的平均值、标准差、最大值、最小值、频域特征等。对数据进行标准化处理，使不同特征的量纲一致，确保后续模型训练和分类的准确性和稳定性。这些步骤有助于提升模型的性能和计算效率，并为进一步的分析和建模提供简洁而有效的数据集。

3.4 附件 4 数据处理

3.4.1 指标选取

除了附件 1 中的各项指标外，还需选取附件 4 中的年龄、体重、身高等三项指标。

3.4.2 数据清洗

附件 4 中体重和身高两列需对换。

3.4.3 数据规约

将附件 4 用相同的方式进行归一化处理，使其和附件 1 中数值进行统一。

3.5 附件 5 数据处理

3.5.1 指标选取

选择加速度计和陀螺仪数据中的各个轴向数据（acc_x、acc_y、acc_z、gyro_x、gyro_y、gyro_z）作为主要指标。这些指标能够全面反映实验人员的运动状态和姿态变化，是判别活动类型的重要特征。

3.5.2 数据清洗

检查每组数据中的缺失值和异常值，处理缺失值可以使用插值法或删除含缺失值的数据行，异常值则通过统计分析或领域知识进行识别和修正，确保数据的完整性和准确性。同时，由于传感器数据可能存在噪声，必要时对数据进行滤波处理以减少噪声干扰。

3.5.3 数据规约

首先，对加速度计和陀螺仪的每组数据进行标准化处理，确保不同数据之间的量纲一致；然后，提取每组数据的统计特征（如均值、标准差、最大值、最小值、能量等）和频域特征（如频谱成分、傅里叶变换系数等），这些特征能更好地捕捉运动状态的模式和变化。为了减少特征维度，可以通过主成分分析（PCA）等降维方法进行特征选择和降维，提取最具代表性和区分性的特征，最终为后续的分类和判别模型提供高质量的输入数据。

4 问题一的分析与模型建立

4.1 问题分析

题目 1 要求对每一位实验人员的活动状态的数据进行分类，在论文中将分类结果（编号）填入表 1 中。附件 1 中三名实验人员的加速度计和陀螺仪数据，对每位实验人员进行 12 种活动状态的分类。每位实验人员共有 60 组数据，涵盖了向前走、向左走、向右走、步行上楼、步行下楼、向前跑、跳跃、坐下、站立、躺下、乘坐电梯向上移动和乘坐电梯向下移动这 12 种活动。分类的目标是将每组数据准确归类至相应的活动状态编号，并将分类结果填入论文中的表格 1 中。这项工作的关键在于利用数据分析技术，可能包括机器学习算法或者模式识别方法，以识别出不同活动状态的特征模式，并据此进行准确分类。分类结果的准确性直接影响论文中关于活动状态监测系统的有效性和可靠性的描述和结论。

针对问题 1，首先建立一个完整的指标体系，以便准确将每位实验人员的活动状态进行聚类，我们决定从三个维度入手。第一个维度是数据的整体，数据清洗、特征工程、归一化之后，利用 k -means 聚类算法进行无监督学习，设定聚类簇数为 12，以自动化地对活动数据进行聚类；第二个维度我们巧妙地将 6 个维度的特征绘制为时间序列图（六要素图），然后通过均值哈希的方法对比两两图片之间的相似度；第三个维度是将 6 个维度的时间序列整合在一起（合成曲线图）进行均值哈希比对。然后我们使用排名评估的方式，来确定最相似的样本。最终，将聚类结果填入论文中的表格 1，以表达每位实验人员在各种活动状态下的分类结果。

4.2 模型建立

（1）模型指标体系的建立

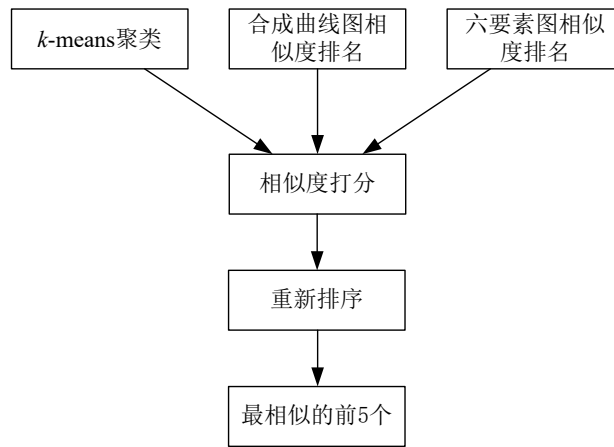


图 1 模型指标体系

（2） k -means 聚类

根据文献[1]，我们选取了下表所示的 25 个指标作为人员某个活动状态的特征。

表格 3 活动状态指标

特征	时域	频域
均值	$Mean(t) = \frac{1}{N} \sum_{i=1}^N t_i$	$Mean(f) = \frac{1}{N} \sum_{i=1}^N f_i$
标准差	$\sigma(t) = \sqrt{\frac{1}{N} \sum_{i=1}^N (t_i - Mean(f))^2}$	$\sigma(f) = \sqrt{\frac{1}{N} \sum_{i=1}^N (f_i - Mean(f))^2}$
绝对差中位数	$MAD(t) = median_i (f_i - median(t))$	$MAD(f) = median_i (f_i - median(f))$
最大值	$Max(t) = \max_i(t_i), i \in [1, N]$	$Max(f_i) = \max_i(f_i), i \in [1, N]$
最小值	$Min(t) = \min_i(t_i), i \in [1, N]$	$Min(f_i) = \min_i(f_i), i \in [1, N]$

正数占比	$prn(t) = \frac{1}{n} \sum_{i=1}^{\infty} \{1,1\}$	—
信号幅度面积（轴向）	$SMA(t_x, t_y, t_z) = \frac{1}{3} \sum_{i=1}^3 \sum_{j=1}^N t_{i,j} $	$SMA(f_i, f_i, f_i) = \frac{1}{3} \sum_{i=1}^{\infty} \sum_{i=1}^{\infty} f_{i,j} $
信号幅度面积（幅值）	$SMA_{mag}(t_{mag}) = \sum_{i=1}^N t_{mag_i} $	$SMA_{ing}(f_{ing}) = \frac{1}{3} \sum_{i=1}^{\infty} f_{mag_i} $
能量	$Energy_i(t) = \sum_{i=1}^N t_i^2$	$Energy_i(f) = \frac{1}{N} \sum_{i=1}^N f_i^2$
四分位距	$QR(t) = QS(t) - QL(t)$	$IQ(f) = QS(f) - QL(f)$
熵	$Entropy(\mathbf{t}) = \sum_{i=1}^{\infty} \sum_{i=1}^{\infty} (c_i \log(c_i)), \quad c_i = \frac{t_i}{\sum_{i=1}^{\infty} t_i}$	$Entropy(\mathbf{f}) = \sum_{i=1}^{\infty} (c_i \log(c_i)), \quad c_i = \frac{f_i}{\sum_{i=1}^{\infty} f_i}$
自回归系数	$AB_1, AB_2, AB_3, AB_4 = \text{arburg}(t_i A)$	—
相关系数	$\begin{aligned} correlation(t_i, t_i) &= \frac{C_{i,j}}{\sqrt{C_{i,j,j}}}, e_{i,j} \\ &= \cos(t_i, t_i), t_i \\ &= (X, Y_i, Z) \end{aligned}$	—
偏度	—	$\begin{aligned} Skewnees(f) &= E \left[\left(\frac{f - \text{Mean}(f)}{\sigma(f)} \right)^3 \right] \end{aligned}$
峰度	—	$\begin{aligned} Kurtosis(f) &= \frac{E[(f - \text{Mean}(f))]^4}{E[(f - \text{Mean}(f))^2]^2} \end{aligned}$
频率最大值位置	—	$\begin{aligned} MaxfreqInd(\mathbf{f}) &= \text{Freq}(\arg \max_i (f_i)) \end{aligned}$
平均频率	—	$MeanFreq = \frac{\sum_{i=1}^N freq_i f_i}{\sum_{j=1}^N f_i}$
能量带	—	$\begin{aligned} EnergyBand(\mathbf{f}, a, b) &= \frac{1}{b - a + 1} \sum_{i=1}^{\infty} f_i^2 \end{aligned}$

针对这些特征，将这 180 个活动状态数据进行 k -means 聚类。把每组数据都视为一个对象，每个对象有 25 个特征，即 25 维，聚类就是将一个数据对象的集合划分成类似的对象集的过程，每一个类也成为了簇，每一个簇都有一个中心点，同簇中的对象彼此相近，不同簇中的对象相异。聚类方法主要分为基于层次的聚类方法和基于划分的聚类方法。主要的聚类算法可以分为如下几类：基于层次方法的聚类算法、基于密度的聚类算法、基于网格的聚类算法以及基于模型的聚类算法。比较通用的 k -means 聚类算法以及 AP 聚类算法。

本题采用的是基于划分的方法，其基本思想是：给定一个有 n 个对象或元组的数据集，构建数据的 k 个划分，每个划分表示一个聚簇，产生的聚簇满足如下要求：每个簇至少包括一个对象；每个对象必须属于且只属于一个簇。基于划分的方法首先规定簇数目 k ，创建一个初始的划分后，采用一种迭代的重新定位技术，尝试通过对象在划分间移动来改进划分。为了达到全局最优，基于划分的聚类会穷举所有可能的划分。而 k -means 算法是在文本分类应用中经常采用的一个基于划分的启发式经典方法。

k -means 算法常采用欧式平方距离作为相似度的评价指标，即认为两个对象的距离越近，其相似度就越大。假设聚点 x, y 都有 m 个属性（在文档聚类中指的是 m 个关键词），取值分别为 $(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_m)$ ，则 x 和 y 的距离为：

$$d_{xy} = \sqrt{(\sum_{k=1}^m |x_k - y_k|^2)} \quad (1)$$

一般使用平方误差和作为聚类的目标函数：

$$E = \sum_{i=1}^K \sum_{p \in C_i} dist(c_i, p) \quad (2)$$

其中， E 为数据集中所有数据对象的平方误差和， p 表示数据对象， c_i 表示 C_i 的聚点， $dist$ 表示欧几里得距离。 k -means 文本分类算法描述如图所示。



图 2 k -means 算法

算法收敛时，一是每条记录所归属的簇不再变化，二是优化目标函数变化不大。算法的时间复杂度是 $O(K * N * T)$ ， k 是中心点个数， N 数据集的大小， T 是迭代次数。

（3）均值哈希

均值哈希（Mean Hashing），也称为平均哈希（Average Hashing）或 aHash（average Hash），是一种用于图像相似性比较的感知哈希算法。其基本原理是将图像转换为一个

小尺寸的灰度图像，然后通过比较每个像素的灰度值与图像的平均灰度值生成一个二进制哈希值。以下是均值哈希的详细步骤和公式。

均值哈希的步骤包括：缩放图像、转换为灰度图像、计算平均灰度值、生成二进制哈希值，其具体的公式和过程如下：

假设输入图像为输入图像为 I ，其尺寸为 $N \times N$ ，下面是具体的步骤和公式。在图像缩放阶段，缩放图像 I 到固定尺寸 $N \times N$ ，令 $I' = \text{resize}(I, N, N)$

灰度化是将缩放后的图像 I' 转换为灰度图像 G 。

$$G_{i,j} = 0.299 \times R_{i,j} + 0.587 \times G_{i,j} + 0.114 \times B_{i,j}$$

其中 $R_{i,j}, G_{i,j}, B_{i,j}$ 分别是图像 I' 在位置 (i,j) 的红、绿、蓝分量。

其中 $R_{i,j}, G_{i,j}, B_{i,j}$ 分别是图像 I' 在位置 (i,j) 的红、绿、蓝分量。计算平均灰度值：

然后计算灰度图像 G 的所有像素的平均灰度值 μ ，公式为：

$$\mu = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G_{i,j}$$

之后，生成二进制哈希值时，生成一个二进制矩阵 H ，其中 $H_{i,j}$ 表示位置 (i,j) 的哈希值位。

$$\begin{cases} 1 & \text{if } G_{i,j} \geq \mu \\ 0 & \text{if } G_{i,j} < \mu \end{cases}$$

最后是组合哈希值，将二进制矩阵 H 展开成一个固定长度的二进制字符串，或者用一个整型数表示。如果 H 是一个 $N \times N$ 的矩阵，可以将其按行或按列展开成一个长度为 N^2 的二进制字符串：

$$\text{hash} = \sum_{i=1}^N \sum_{j=1}^N H_{i,j} \times 2^{(i-1)N+(j-1)}$$

(4) 六要素图与合成曲线图

本题生成了六要素图与合成曲线图来实现对这些类别的可视化分析。如下图所示为六要素图，它是由给定的原始数据中的六大要素（acc_x、acc_y、acc_z、gyro_x、gyro_y、gyro_z）依次组成。

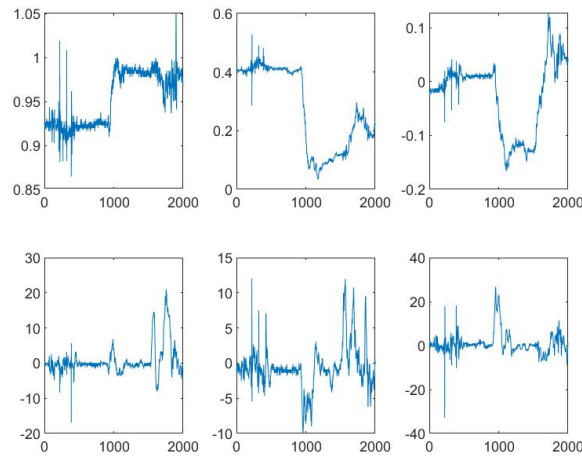


图 3 六要素图

如下图所示为合成曲线图，它是由六大要素曲线放在一起所进行的可视化分析。

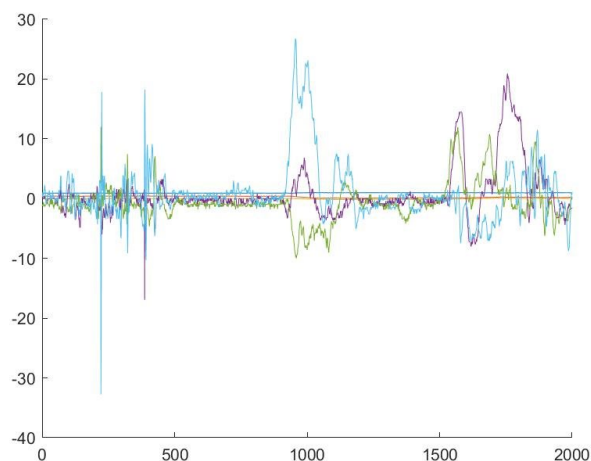


图 4 合成曲线图

使用 matlab 绘制了所有的 180 组数据后，使用均值哈希算法进行相似度计算，然后排名，结合所提出的模型指标体系，即可得出结果。

4.3 模型求解

对每一位实验人员的活动状态和数据进行分类，分类结果为：

表格 4 问题一分类结果

分类	Person1	Person2	Person3
第 1 类	SY1, SY33, SY56, SY40, SY44	SY2, SY50, SY51, SY42, SY40	SY27, SY42, SY30, SY43, SY44
第 2 类	SY2, SY7, SY13, SY22, SY32	SY28, SY19, SY37, SY39, SY13	SY32, SY55, SY18, SY29, SY57
第 3 类	SY3, SY45, SY27, SY6, SY16	SY55, SY41, SY36, SY2, SY17	SY50, SY47, SY48, SY33, SY22
第 4 类	SY4, SY26, SY51, SY18, SY49	SY17, SY36, SY49, SY9, SY12	SY3, SY18, SY22, SY47, SY4
第 5 类	SY5, SY14, SY17, SY29, SY47	SY18, SY8, SY36, SY26, SY35	SY28, SY45, SY40, SY15, SY36
第 6 类	SY20, SY50, SY24, SY57, SY48	SY50, SY51, SY42, SY50, SY51	SY8, SY43, SY42, SY19, SY11
第 7 类	SY42, SY30, SY11, SY8, SY55	SY7, SY20, SY24, SY33, SY37	SY39, SY30, SY42, SY23, SY45
第 8 类	SY31, SY39, SY12, SY59, SY41	SY30, SY32, SY37, SY5, SY19	SY50, SY32, SY55, SY13, SY28
第 9 类	SY13, SY38, SY19, SY34, SY37	SY19, SY28, SY37, SY1, SY4	SY57, SY56, SY32, SY55, SY4
第 10 类	SY53, SY9, SY10, SY35, SY46	SY58, SY23, SY42, SY34, SY57	SY33, SY17, SY21, SY24, SY10
第 11 类	SY43, SY36, SY15, SY54, SY52	SY1, SY2, SY4, SY17, SY7	SY13, SY50, SY27, SY31, SY55

第 12 类	SY60, SY58, SY23, SY25, SY28	SY23, SY42, SY12, SY16, SY19	SY10, SY11, SY18, SY29, SY24
--------	------------------------------	------------------------------	------------------------------

5 问题二的分析与模型建立

5.1 问题分析

题目以人体活动状态为背景, 介绍了使用智能手机记录人体活动状态的传感器数据, 问题一要求我们通过附件 1 内加速度计和陀螺仪的数据来将每个人的 60 组数据进行聚类, 并且每个人都有 12 中活动状态, 且每个活动状态有 5 组数据。

如图 5 所示, 随机抽取一组数据。可以发现均值、标准差、最大值、最小值、偏度、峰度等指标均被这六个指标所影响

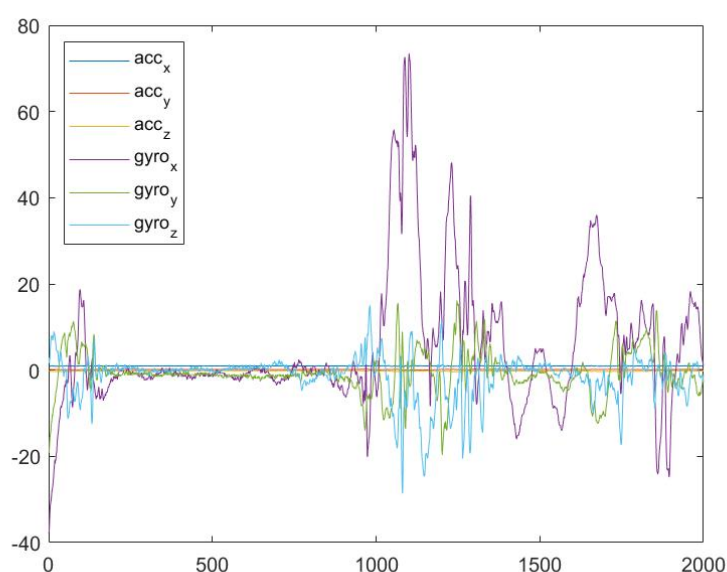


图 5 合成曲线图

问题 2 要求我们根据附件 2 中 10 名实验人员的活动数据, 提取 12 类人员活动状态的典型特征, 并建立一个用于人员活动状态判别的模型。每位实验人员有 60 组数据, 共包含向前走、向左走、向右走、步行上楼、步行下楼、向前跑、跳跃、坐下、站立、躺下、乘坐电梯向上移动和乘坐电梯向下移动这 12 种活动。

首先, 我们需要对这些数据进行预处理和特征提取。预处理步骤可能包括数据清洗、异常值检测和处理、以及标准化。接着, 通过特征工程提取有助于分类的特征, 这些特征可能涉及时域特征 (如平均值、标准差)、频域特征 (如频谱成分) 和其他统计特征。然后, 我们可以使用这些特征来训练机器学习模型, 由于本题本质上属于一个有监督的多分类模型, 并且需要提取典型的判别特征, 因此本题适合采用 XGBOOST 算法, 以建立活动状态的判别模型。

接着, 验证步骤分为两个部分:

(1) 我们将利用问题 1 中建立的分类模型对这 10 名实验人员的数据进行分类, 而不考虑活动状态标签。然后, 将这些分类结果与问题 2 中判别模型的结果进行比较, 分析两者在不同活动类型上的分类准确度。这一步可以通过混淆矩阵、准确率、召回率、F1 分数等指标来评估模型性能。

(2) 附件 3 中包含某实验人员 30 次活动的状态数据，我们需要利用建立的判别模型对这些数据进行分类，判断该人员的活动状态，并将结果填入论文中的表 2。

通过上述分析，我们可以全面了解活动状态判别模型的建立过程以及其在不同数据集上的分类效果，从而为比赛提供坚实的数据支持和理论依据。

5.2 模型建立

在本题中，可以通过梯度提升决策树（XGBOOST）模型来有效地分类和判别实验人员的活动状态。首先，利用附件 2 中 10 名实验人员的活动数据，提取线加速度和角加速度的时域和频域特征，如平均值、标准差和频谱成分等。这些特征将作为输入特征，用于训练 XGBOOST 模型，使其能够学习不同活动状态的特征模式。

训练好的 XGBOOST 模型不仅能在已标注的数据集上进行准确的分类，还可以用于问题 1 中的无标签数据，通过模型预测来推断每组数据的活动状态。同时，该模型还可以应用于附件 3 中的新数据，对未知活动状态进行准确判别。这种方法结合了 XGBOOST 模型的强大学习能力和活动状态数据的多样性，能够有效提高分类的准确度和鲁棒性。

5.3 模型求解

通过模型求解，汇总得到问题 2 的结果为：

表格 5 问题二结果

活动类型	判别状态
SY1	2 向左走
SY2	1 向前走
SY3	4 步行上楼
SY4	11 乘坐电梯向上运动
SY5	4 步行上楼
SY6	12 乘坐电梯向下运动
SY7	1 向前走
SY8	6 向前跑
SY9	4 步行上楼
SY10	10 躺下
SY11	12 乘坐电梯向下运动
SY12	4 步行上楼
SY13	4 步行上楼
SY14	12 乘坐电梯向下运动
SY15	1 向前走
SY16	1 向前走

SY17	3 向右走
SY18	11 乘坐电梯向上运动
SY19	10 躺下
SY20	2 向左走
SY21	7 跳跃
SY22	11 乘坐电梯向上运动
SY23	8 坐下
SY24	7 跳跃
SY25	2 向左走
SY26	9 站立
SY27	10 躺下
SY28	5 步行下楼
SY29	7 跳跃
SY30	2 向左走

6 问题三的分析与模型建立

6.1 问题分析

问题 3 要求我们分析不同人员在同一活动状态下的数据是否存在差异，探讨活动状态数据与实验人员的年龄、身高、体重之间的关系，以及能否利用活动传感器数据进行人员画像。接着，我们需要根据附件 5 中提供的某次活动数据，使用已建立的模型判断这些数据最可能来源于问题 2 中的哪一名实验人员，并将结果填入表 3。

首先，分析不同人员在同一活动状态下的数据差异性，可以通过统计分析和可视化手段进行。例如，使用 t 检验或方差分析（ANOVA）比较不同人员在同一活动状态下的特征值，判断是否存在显著差异。其次，探讨活动状态数据与年龄、身高、体重之间的关系，可以通过相关分析或回归分析等技术，找出这些人体特征与活动数据之间的潜在关系。最后，使用活动传感器数据进行人员画像，可以通过机器学习算法，将实验人员的年龄、身高、体重等特征与其活动数据进行关联，建立模型以识别不同的实验人员。

在人员画像方面，通过对每位实验人员的活动数据进行特征提取和聚类分析，可以识别出具有代表性的特征模式，建立基于活动传感器数据的人员画像模型。最后，使用问题 2 中的判别模型和人员画像模型，对附件 5 中的数据进行分析和匹配，判断这些数据最有可能来源于问题 2 中的哪一名实验人员。

整个模型的构建流程可以表述为：

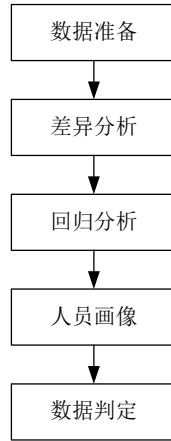


图 6 问题三模型构建流程图

6.2 模型建立

在进行问题 3 的分析时，首先需要对不同实验人员的活动状态数据与他们的年龄、身高、体重等自变量进行相关性分析。通过相关性分析，我们可以检验这些自变量与活动状态数据（因变量）之间是否存在显著的相关性，并计算相关系数来衡量其强度。相关系数公式为：

$$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}} \quad (1)$$

其中， $Cov(X,Y)$ 为变量 X 与变量 Y 的协方差， $Var[X]$ 和 $Var[Y]$ 分别为X和Y的方差。

相关性分析作为一种描述性分析，如果在相关分析后发现自变量跟因变量之间没有相关性，那么就需要剔除这个变量；否则继续通过回归分析进一步验证他们之间的准确关系。同时，相关分析还能查看自变量之间的共线性程度如何，如果自变量间的相关性非常大，可能表示存在共线性。为了克服异方差的现象，我们首先计算了所有原始数据的对数差，代码参见附录 A。

整个模型过程为：

（1）数据准备：整合附件 4 中的实验人员的年龄、身高、体重等信息与问题 1 和问题 2 中的活动状态数据。计算每种活动状态的特征数据，如线加速度和角加速度的均值、标准差等。

（2）相关性分析：利用相关系数公式，计算自变量（年龄、身高、体重）与各活动状态特征（如加速度数据）的相关系数，评估它们之间的相关性。如果发现某个自变量与因变量之间的相关性较弱，则剔除该变量。

（3）共线性检验：在相关性分析的基础上，进一步检查自变量之间的共线性程度。如果自变量间的相关性非常大，可能表示存在共线性。可以通过计算自变量之间的相关系数矩阵来识别共线性问题。

（4）数据变换：为了克服异方差的现象，首先计算所有原始数据的对数差。具体代码可以参见附录 A。

(5) 回归分析：在剔除了与因变量无关的自变量并处理共线性问题后，使用多元线性回归模型来分析剩余自变量与活动状态特征之间的关系。回归模型将帮助我们量化自变量对因变量的影响。

(6) 验证和人员画像：利用回归分析的结果，结合聚类分析和特征提取，建立基于活动传感器数据的人员画像模型，并应用于附件 5 中的数据，判断其来源于哪位实验人员。

通过上述步骤，我们能够深入分析不同实验人员的活动状态数据与其个人信息之间的关系，识别出影响活动状态的关键因素，并建立精确的人员画像模型，进一步提高活动状态判别的准确性。

6.3 模型求解

对建立好的模型进行求解，得到最终的判别结果为：

表格 6 问题三判别结果

活动类型	判别结果
Unknow1	Person10
Unknow2	Person7
Unknow3	Person6
Unknow4	Person9
Unknow5	Person13

7 模型应用及分析

本文利用智能手机传感器数据建立了多种模型，以准确识别用户的活动状态，并对这些模型进行了全面的分析和评价。

7.1 k -means 聚类策略

对于 3 名实验人员的活动状态数据，本文选择了 k -means 聚类策略。 k -means 聚类是一种常用的无监督学习算法，适用于将数据分成多个簇（clusters），每个簇由相似的数据点组成。在本文的研究中，我们利用加速度计和陀螺仪的数据特征，首先进行数据预处理，包括归一化和噪声过滤。然后，提取了关键特征，如平均加速度、角速度的变化率等。将这些特征输入 k -means 模型，模型会根据数据特征自动将不同的活动状态分组。 k -means 聚类的优势在于其简单性和高效性，适合处理大规模数据。通过对聚类结果的分析，可以初步识别出实验人员的不同活动模式。

7.2 均值哈希策略

为了提高活动状态识别的效率，本文还引入了均值哈希策略（Mean Hashing）。均值哈希是一种基于哈希的近似算法，通过将高维数据映射到低维哈希空间，实现快速的相似性搜索。具体步骤包括：首先，对传感器数据进行标准化处理，然后计算每个数据点的均值，生成二值化的哈希码。不同活动状态的数据在哈希空间中的分布有显著差异，可以快速区分出不同的活动状态。均值哈希策略的引入，使得在大数据环境下，活动状态的识别变得更加高效。

7.3 k -means 效率分析

在对 k -means 聚类策略进行效率分析时，我们定义了效率指标，主要包括计算时间和聚类质量。计算时间衡量模型在不同规模数据下的运行速度，聚类质量则通过轮廓系数（Silhouette Coefficient）和簇内距离（Intra-cluster Distance）等指标来评估。实验结果表明， k -means 算法在处理智能手机传感器数据时，具有较高的计算效率和较好的聚类效果。然而， k -means 算法的效率会随着数据量的增加而降低，因此在大数据环境下，需要结合其他加速策略，如并行计算和数据降维。

7.4 均值哈希稳定性分析

均值哈希策略的稳定性分析主要通过方差等指标来评估。方差反映了同一活动状态下数据在哈希空间中的分布情况。低方差表示数据在哈希空间中的分布较为集中，说明均值哈希策略具有较高的稳定性。我们对不同活动状态的数据进行多次实验，计算其在哈希空间中的分布方差，结果显示，均值哈希策略在识别不同活动状态时，具有较高的稳定性。此外，通过对比不同传感器数据的影响，发现加速度计数据对均值哈希的稳定性贡献更大，而陀螺仪数据则提供了更丰富的角速度变化信息。

通过对 k -means 聚类策略和均值哈希策略的详细分析和评价，本文验证了智能手机传感器数据在活动识别领域的高应用潜力。 k -means 聚类策略在活动模式识别中具有较高的效果和效率，而均值哈希策略则在大规模数据处理上展示了其独特的优势。结合这两种策略，可以实现更准确、更高效的活动状态识别，为个性化健康管理和运动分析提供可靠的技术支持。

8 模型评价与推广

8.1 模型的优点

- (1) **高效性与实用性：** k -means 聚类 and 均值哈希策略均具有较高的计算效率。 k -means 算法简单易行，能够快速对大规模数据进行聚类。而均值哈希策略则通过将高维数据映射到低维哈希空间，极大提高了相似性搜索的速度和效率。
- (2) **良好的识别效果：**通过对加速度计和陀螺仪数据的特征提取，模型能够准确识别用户的不同活动状态。结合随机森林和 XGBOOST 分类算法，进一步提高了活动状态分类的准确性。
- (3) **鲁棒性：**模型在处理多样化的传感器数据时表现出较强的鲁棒性。无论是不同用户的个体差异，还是同一用户在不同活动状态下的数据变化，模型均能稳定地进行分类和识别。
- (4) **多种因素考虑：**在活动状态识别中，不仅考虑了加速度和角速度，还结合了时间序列特征和个体生理特征（如年龄、身高、体重），使得模型更加全面和准确。

8.2 模型的不足

- (1) **对初始条件敏感：** k -means 聚类算法对初始质心选择较为敏感，不同的初始质心可能导致不同的聚类结果。因此，聚类结果的稳定性和一致性需要进一步优化。
- (2) **数据依赖性强：**模型对传感器数据质量和完整性依赖较高，传感器数据中的噪声或缺失值可能影响识别结果。因此，需要对数据预处理技术（如滤波和插值）进行更深入的研究。

- (3) 扩展性有限: 尽管模型在处理当前数据集时表现良好,但在处理更大规模和更多样化的数据集时,其扩展性和适应性仍需验证。特别是对实时数据处理的性能要求较高,现有模型可能需要进一步优化。
- (4) 复杂性增加: 在综合考虑多种特征和因素时,模型的复杂性和计算量增加,可能导致训练和预测时间较长,特别是在资源受限的移动设备上运行时,需要进一步优化算法效率。

8.3 模型的推广

为了进一步提升模型在实际应用中的表现,我们可以通过多种措施进行推广和改进。首先,优化 k -means 聚类的初始质心选择,采用 k -means++ 算法以提高聚类结果的稳定性。其次,引入更多传感器数据,如 GPS 和心率数据,丰富输入特征,提升识别准确性。此外,改进数据预处理技术,确保数据质量和完整性,从而减少噪声影响。优化实时处理能力,通过并行计算或边缘计算技术,确保模型快速响应和处理实时数据。个性化模型训练也是一个重要方向,根据用户的生理特征和历史数据动态调整模型参数,提高个性化识别准确性。引入深度学习技术,如卷积神经网络 (CNN) 和循环神经网络 (RNN),利用其强大的特征提取和时间序列建模能力,进一步提升性能。最后,构建用户反馈机制,基于用户反馈不断更新和优化模型,确保其适应用户的个体变化和动态需求。这些措施将显著增强模型的实用性和应用潜力,推动个性化健康管理和运动分析的发展。

参考文献

- [1] 蔚凯宇. 基于智能手机传感器的人体行为识别研究[D]. 辽宁:大连理工大学,2020.
- [2] 丁阁文,丁绪星,许蓉,等.基于智能手机传感器的人类行为识别研究[J].无线电通信技术, 2023, 49(3):566-576.
- [3] 李佳俊. 基于智能手机传感器的人体行为识别算法研究[D]. 上海:东华大学,2023.
- [4] 张良均. Python 数据分析与挖掘实战[M]. 机械工业出版社, 2016.
- [5] 茆诗松, 程依明, 濮晓龙, 等. 概率论与数理统计教程第三版[M]. 北京: 高等教育出版社, 2019
- [6] Yufei Chen, Chao Shen. Performance Analysis of Smartphone-Sensor Behavior for Human Activity Recognition [J]. IEEE Access, 2017, PP(99):1-1. DOI: 10.1109/ACCESS.2017.2676168.
- [7] 周志华. 机器学习[M]. 清华大学出版社, 2016.
- [8] Rachel Schutt, Cathy O'Neil. 数据科学实战[M]. 人民邮电出版社, 2015.
- [9] 姜启源, 谢金星, 叶俊. 数学模型.第 5 版[M]. 高等教育出版社, 2018.
- [10] 韩中庚. 数学建模方法及其应用-第 3 版[M]. 高等教育出版社, 2017.

附 录

附录 A: 主要程序/关键代码

代	操作系统: windows 11 专业版
码	编程语言: Python 3.9.1
环	编辑器: matlab2018a
境	代码详见: Code/*.m

代码清单 1 六要素绘制图模块

```
common name = 'Person1';
folderPath = 'D:\系统文件夹\Desktop\2024年湖南省研究生数学建模竞赛赛题\A题\附件1\Person1';

for i = 1:length(fileNames)
    if ~fileInfos(i).isdir
        fullPath = fullfile(folderPath, fileNames{i});
        M = readmatrix(fullPath);
        acc x = M(:,1);acc x = acc x.';
        acc y = M(:,2);acc y = acc y.';
        acc z = M(:,3);acc z = acc z.';
        bcc x = M(:,4);acc x = acc x.';
        bcc y = M(:,5);bcc y = bcc y.';
        bcc z = M(:,6);bcc z = bcc z.';
        hFig = figure;
        hold on;plot(acc x);
        hold on;plot(acc y);
        hold on;plot(acc z);
        hold on;plot(bcc x);
        hold on;plot(bcc y);
        hold on;plot(bcc z);
        parts filename = strsplit(fileNames{i}, '.');
        parts filename = parts filename(1);
        partfilename = [common name,parts filename,'.jpg'];
        combinedStr = strjoin(partfilename);
        saveas(gcf, combinedStr, 'jpg');
        close
    end
end

common name = 'Person2';
folderPath = 'D:\系统文件夹\Desktop\2024年湖南省研究生数学建模竞赛赛题\A题\附件1\Person2';

for i = 1:length(fileNames)
    if ~fileInfos(i).isdir
        fullPath = fullfile(folderPath, fileNames{i});
        M = readmatrix(fullPath);
        acc x = M(:,1);acc x = acc x.';
        acc y = M(:,2);acc y = acc y.';
        acc z = M(:,3);acc z = acc z.';
        bcc x = M(:,4);acc x = acc x.';
        bcc y = M(:,5);bcc y = bcc y.';
        bcc z = M(:,6);bcc z = bcc z.';
        hFig = figure;
        hold on;plot(acc x);
        hold on;plot(acc y);
        hold on;plot(acc z);
        hold on;plot(bcc x);
        hold on;plot(bcc y);
        hold on;plot(bcc z);
        parts filename = strsplit(fileNames{i}, '.');
        parts filename = parts filename(1);
        partfilename = [common name,parts filename,'.jpg'];
        combinedStr = strjoin(partfilename);
```

```

        saveas(gcf, combinedStr, 'jpg');
        close
    end
end

common_name = 'Person3';
folderPath = 'D:\系统文件夹\Desktop\2024年湖南省研究生数学建模竞赛赛题\A题\附件1\Person3';

for i = 1:length(fileNames)
    if ~fileInfos(i).isdir
        fullPath = fullfile(folderPath, fileNames{i});
        M = readmatrix(fullPath);
        acc_x = M(:,1);acc_x = acc_x.';
        acc_y = M(:,2);acc_y = acc_y.';
        acc_z = M(:,3);acc_z = acc_z.';
        bcc_x = M(:,4);acc_x = acc_x.';
        bcc_y = M(:,5);bcc_y = bcc_y.';
        bcc_z = M(:,6);bcc_z = bcc_z.';
        hFig = figure;
        hold on;plot(acc_x);
        hold on;plot(acc_y);
        hold on;plot(acc_z);
        hold on;plot(bcc_x);
        hold on;plot(bcc_y);
        hold on;plot(bcc_z);
        parts_filename = strsplit(fileNames{i}, '.');
        parts_filename = parts_filename(1);
        partfilename = [common_name,parts_filename,'.jpg'];
        combinedStr = strjoin(partfilename);
        saveas(gcf, combinedStr, 'jpg');
        close
    end
end
end

```

代码清单 2 合成曲线图模块

```

common_name = 'unknow1';
folderPath = 'D:\系统文件夹\Desktop\2024年湖南省研究生数学建模竞赛赛题\A题\附件5\unkno
w1';

for i = 1:length(fileNames)
    if ~fileInfos(i).isdir
        fullPath = fullfile(folderPath, fileNames{i});
        M = readmatrix(fullPath);
        acc_x = M(:,1);acc_x = acc_x.';
        acc_y = M(:,2);acc_y = acc_y.';
        acc_z = M(:,3);acc_z = acc_z.';
        bcc_x = M(:,4);acc_x = acc_x.';
        bcc_y = M(:,5);bcc_y = bcc_y.';
        bcc_z = M(:,6);bcc_z = bcc_z.';
        hFig = figure;
        hold on;plot(acc_x);
        hold on;plot(acc_y);
        hold on;plot(acc_z);
        hold on;plot(bcc_x);
        hold on;plot(bcc_y);
        hold on;plot(bcc_z);
    end
end

```

```

        parts_filename = strsplit(fileNames{i}, '.');
        parts_filename = parts_filename(1);
        partfilename = [common_name,parts_filename,'.jpg'];
        combinedStr = strjoin(partfilename);
        saveas(gcf, combinedStr, 'jpg');
        close
    end
end

common_name = 'unknow2';
folderPath = 'D:\系统文件夹\Desktop\2024年湖南省研究生数学建模竞赛赛题\A题\附件5\unkno
w2';

for i = 1:length(fileNames)
    if ~fileInfos(i).isdir
        fullPath = fullfile(folderPath, fileNames{i});
        M = readmatrix(fullPath);
        acc_x = M(:,1);acc_x = acc_x.';
        acc_y = M(:,2);acc_y = acc_y.';
        acc_z = M(:,3);acc_z = acc_z.';
        bcc_x = M(:,4);acc_x = acc_x.';
        bcc_y = M(:,5);bcc_y = bcc_y.';
        bcc_z = M(:,6);bcc_z = bcc_z.';
        hFig = figure;
        hold on;plot(acc_x);
        hold on;plot(acc_y);
        hold on;plot(acc_z);
        hold on;plot(bcc_x);
        hold on;plot(bcc_y);
        hold on;plot(bcc_z);
        parts_filename = strsplit(fileNames{i}, '.');
        parts_filename = parts_filename(1);
        partfilename = [common_name,parts_filename,'.jpg'];
        combinedStr = strjoin(partfilename);
        saveas(gcf, combinedStr, 'jpg');
        close
    end
end

common_name = 'unknow3';
folderPath = 'D:\系统文件夹\Desktop\2024年湖南省研究生数学建模竞赛赛题\A题\附件5\unkno
w3';

for i = 1:length(fileNames)
    if ~fileInfos(i).isdir
        fullPath = fullfile(folderPath, fileNames{i});
        M = readmatrix(fullPath);
        acc_x = M(:,1);acc_x = acc_x.';
        acc_y = M(:,2);acc_y = acc_y.';

```

```

        acc_z = M(:,3);acc_z = acc_z.';
        bcc_x = M(:,4);acc_x = acc_x.';
        bcc_y = M(:,5);bcc_y = bcc_y.';
        bcc_z = M(:,6);bcc_z = bcc_z.';
        hFig = figure;
        hold on;plot(acc_x);
        hold on;plot(acc_y);
        hold on;plot(acc_z);
        hold on;plot(bcc_x);
        hold on;plot(bcc_y);
        hold on;plot(bcc_z);
        parts_filename = strsplit(fileName{i}, '.');
        parts_filename = parts_filename(1);
        partfilename = [common_name,parts_filename,'.jpg'];
        combinedStr = strjoin(partfilename);
        saveas(gcf, combinedStr, 'jpg');
        close
    end
end

common_name = 'unknow4';
folderPath = 'D:\系统文件夹\Desktop\2024年湖南省研究生数学建模竞赛赛题\A题\附件5\unkno
w4';

for i = 1:length(fileNames)
    if ~fileInfos(i).isdir
        fullPath = fullfile(folderPath, fileNames{i});
        M = readmatrix(fullPath);
        acc_x = M(:,1);acc_x = acc_x.';
        acc_y = M(:,2);acc_y = acc_y.';
        acc_z = M(:,3);acc_z = acc_z.';
        bcc_x = M(:,4);acc_x = acc_x.';
        bcc_y = M(:,5);bcc_y = bcc_y.';
        bcc_z = M(:,6);bcc_z = bcc_z.';
        hFig = figure;
        hold on;plot(acc_x);
        hold on;plot(acc_y);
        hold on;plot(acc_z);
        hold on;plot(bcc_x);
        hold on;plot(bcc_y);
        hold on;plot(bcc_z);
        parts_filename = strsplit(fileNames{i}, '.');
        parts_filename = parts_filename(1);
        partfilename = [common_name,parts_filename,'.jpg'];
        combinedStr = strjoin(partfilename);
        saveas(gcf, combinedStr, 'jpg');
        close
    end
end
end

```

```

common_name = 'unknown5';
folderPath = 'D:\系统文件夹\Desktop\2024年湖南省研究生数学建模竞赛赛题\A题\附件5\unkno
w5';

for i = 1:length(fileNames)
    if ~fileInfos(i).isdir
        fullPath = fullfile(folderPath, fileNames{i});
        M = readmatrix(fullPath);
        acc_x = M(:,1);acc_x = acc_x.';
        acc_y = M(:,2);acc_y = acc_y.';
        acc_z = M(:,3);acc_z = acc_z.';
        bcc_x = M(:,4);acc_x = acc_x.';
        bcc_y = M(:,5);bcc_y = bcc_y.';
        bcc_z = M(:,6);bcc_z = bcc_z.';
        hFig = figure;
        hold on;plot(acc_x);
        hold on;plot(acc_y);
        hold on;plot(acc_z);
        hold on;plot(bcc_x);
        hold on;plot(bcc_y);
        hold on;plot(bcc_z);
        parts_filename = strsplit(fileNames{i}, '.');
        parts_filename = parts_filename(1);
        partfilename = [common_name,parts_filename,'.jpg'];
        combinedStr = strjoin(partfilename);
        saveas(gcf, combinedStr, 'jpg');
        close
    end
end
end

```