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。

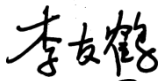
我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

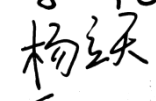
我们授权湖南省研究生数学建模竞赛组委会，可将我们的论文以任何形式进行公开展示（包括进行网上公示，在书籍、期刊和其他媒体进行正式或非正式发表等）。

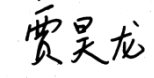
我们参赛选择的题号是（从组委会提供的赛题中选择一项填写）：　　A

我们的参赛编号（请填写完整参赛编号）：202418001012

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. 

2. 

3. 

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

第九届湖南省研究生数学建模竞赛

题　目：使用智能手机记录人体活动状态

摘要

随着人们对健康和体育运动的重视程度不断提高，通过智能手机记录和分析人体活动状态的需求也日益增长。本文针对利用加速度计和陀螺仪中数据分析人体活动这一问题，基于聚类和判别方法，通过确定合加速度平均分段方差和旋转角度等指标，利用 K-means 对无标签数据进行无监督分析；利用有监督机器学习方法构建分类器模型，用以进行运动状态判别及运动人员的判断。

针对问题一：对于 3 名实验人员的运动数据进行分析，聚类不同活动的关键在于特征和模型的选取。综合考虑加速度计和陀螺仪中数据，本文建立了合加速度平均分段方差和旋转角度等指标，将聚类活动状态的问题转化为划分特征值区间的问题，随后基于聚类思想，用 K-means 方法对无标签数据进行了分类。该模型充分考虑了位移和旋转的影响。

针对问题二：对于 10 名实验人员的活动数据进行分析，判别活动状态的关键在于典型特征的选取。综合考虑位移和旋转，本文提取了加速度计和陀螺仪时域和频域的特征，以及三轴之间的皮尔逊相关系数等。基于判别思想，建立了人员活动状态的有监督分类器模型。使用支持向量机、朴素贝叶斯等算法，训练了不同的分类器。比较了问题 2 中判别模型和问题 1 的分类模型，分析了采用分类模型对不同活动类型分类时的分类准确度的下降问题。并给出了附件三中人员的活动状态。

针对问题三：对于 13 位实验人员的年龄、身高、体重等数据进行分析，改进问题二中算法，得出不同人员的同一活动状态存在差异。活动状态数据与实验人员的年龄、身高、体重有关系，能使用活动传感器数据进行人员画像。并且，进一步根据问题 2 的 10 位实验人员中的 5 位的某次活动数据，利用模型判断他们分别最可能来源于问题 2 中对应的实验人员编号。

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，能合理解决提出的问题，具有实用性强，算法效率高等特点，该模型在健康监测、体育训练、智能穿戴等方面也能使用。

关键词：机器学习　K-means　线性分类器　人员画像

目录

1 绪论

1.1 研究背景

随着科技的进步，智能手机内置的传感器种类和性能不断提升。加速度计和陀螺仪作为智能手机中常见的运动传感器，能够分别感知手机的线性加速度和角速度变化。这些传感器通过感知手机在运动过程中的物理量，为分析人体活动提供了丰富的数据基础[1][2]。

智能手机的普及使得人们几乎随时随地携带这一设备，从而能够持续、实时地收集用户的运动数据。这些数据包括步数、跑步速度、行走距离、方向变化等，对于分析人体活动模式具有重要意义。

随着人们对健康和体育运动的重视程度不断提高，通过智能手机记录和分析人体活动状态的需求也日益增长。这些数据可以帮助用户了解自己的运动情况，制定合理的运动计划，从而提高运动效果，预防运动损伤。

随着算法和数据分析技术的不断进步，人们能够从智能手机传感器收集的海量数据中提取出有价值的信息。通过合适的算法和机器学习方法，可以对用户的活动进行分类标记，识别出不同的运动模式，如步行、跑步、骑车等。

基于智能手机传感器的人体活动识别技术具有广泛的应用场景[3]。例如，在健康监测方面，可以实时监测用户的步数、心率等生理指标，评估用户的健康状况[4]；在体育训练方面，可以分析用户的运动姿势、力量分布等，帮助用户优化训练方法；在智能穿戴设备中，可以将手机传感器数据与其他设备（如智能手表、智能手环）的数据进行融合，提供更全面的运动分析[5]。

图 1 基于智能手机传感器的应用

尽管基于智能手机传感器的人体活动识别技术[6]具有广泛的应用前景[7]，但仍面临一些挑战。例如，传感器数据的采集会消耗手机电量，需要设计高效的数据采集和处理方法；算法的精确度和实时性也是一个挑战，需要不断优化算法以提高识别效果。未来，随着传感器技术的进一步发展和算法的不断优化，相信这一技术将在人体活动识别领域发挥更大的作用。

1.2 国内外研究现状

2019 年，美国加利福尼亚州洛杉矶，南加州大学，凯克医学院预防医学系 KenanLi 等人，应用多元分割方法，创建了一个适应可变持续时间活动发作的人类活动识别框架，通过检测多元时间序列中活动发作的变化点和预测由这些变化点定义的每个同质窗口的活动[8]。在活动持续时间可变的场景中，成功从可穿戴传感器数据中识别出人类活动。并且，同以往方法相比，得到的结果更为准确。

具体方法如下：应用标准固定宽度滑动窗口（4-6 种不同大小）和贪婪高斯分割来识别过滤后的三轴加速度计和陀螺仪数据中的断点。在进行标准特征工程后，应用了 Xgboost 模型来预测每个窗口内的身体活动，然后将窗口预测转换为瞬时预测，以便于跨分割方法进行比较。在 2 个数据集中应用了这些方法：数据集 1 为使用智能手机进行人类活动识别数据集，其中共有 30 名成年人在佩戴腰带智能手机时执行了持续时间大致相等的活动（每次约 20 秒）；数据集 2 为儿童哮喘生物医学实时健康评估数据集，其中共有 14 名儿童在佩戴智能手表时执行了 6 项活动，每项活动约 10 分钟。为了模拟真实场景，在儿童哮喘生物医学实时健康评估数据中生成了人为的不等活动持续时间，方法是将每个活动持续时间随机细分为 10 个部分，并随机连接 60 个活动持续时间。每个数据集被分为~90%的训练和~10%的保留测试。

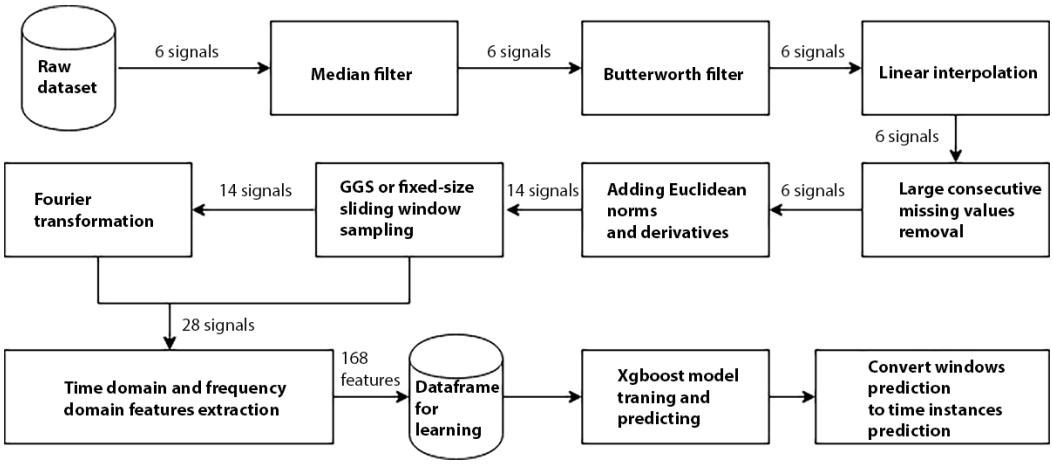


图 2 人类活动识别框架的工作流程

2021 年，香港，香港城市大学数据科学学院，Yu-ChengHsu 等人，基于传感器提出了一种省时且无需耗费大量人力来识别和预处理惯性信号的方法，利用从惯性传感器收集的信号数据开发一个自动运动数据分析框架，用于社区老年人的平衡活动分析[9]。从而为医疗专业人员提供了一种高效的平衡评估工具。从长远来看，该方法可能提供一种灵活的解决方案，以减轻社区持续健康监测的负担。

具体方法如下：研究共招募了 59 名社区老年人（19 名男性和 40 名女性；平均年龄为 81.86 岁，标准差 6.95 岁）。使用随身佩戴的惯性测量装置（包括加速度计和陀螺仪）在每个人的 L4 椎骨处收集数据。在通过卷积长短期记忆(LSTM)神经网络进行数据预处理和运动检测后，采用一类支持向量机(SVM)、线性判别分析(LDA)和 k 最近邻(k-NN)对高风险个体进行分类。利用该研究开发的框架在检测从坐到站、旋转 360°

和从站到坐的动作方面的平均准确率分别为 87%、86%和 89%。使用 Tinetti 性能导向移动性评估-平衡(POMA-B)标准，通过一类 SVM 和 k-NN 对平衡评估分类。在对异常坐立、360°旋转和站立坐立动作进行分类时，准确率分别为 90%、92%和 86%。

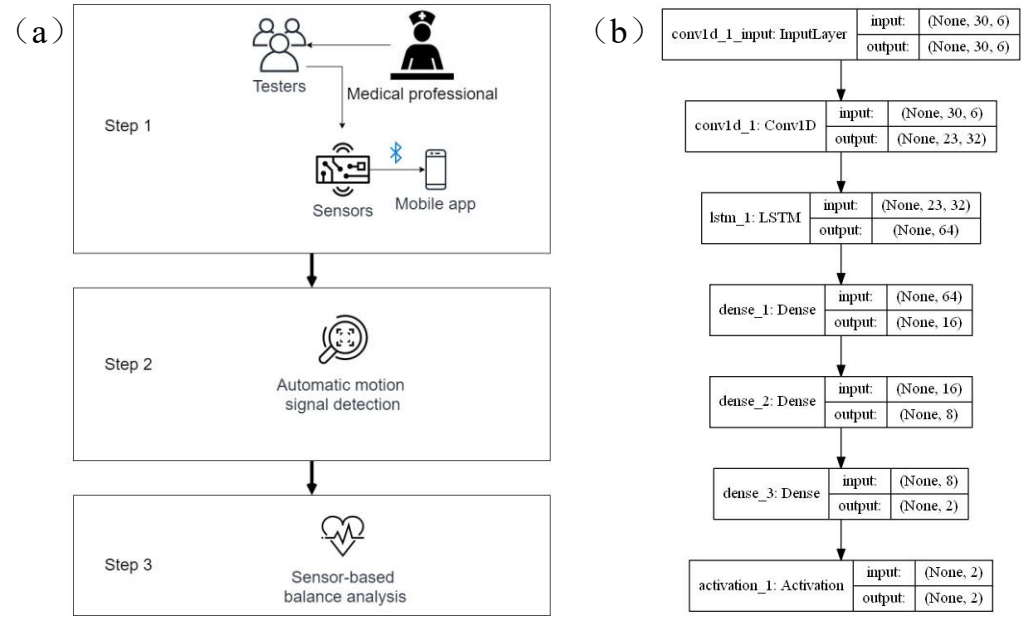


图 3 （a）已开发框架结构示意图（b）用于运动检测的卷积 LSTM 的结构

1.3 基本概念

通过加速度计和陀螺仪的数据，智能手机处理器可以通过感知到手机的姿态、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。采用该方法的优势是可以减少噪音影响，提供补充信息，确保快速响应率。但是，该方法存在一定的局限性，无法实时检测姿势以及姿势与动作之间的关联。因此，需要本文后期采用算法等进一步判断具体活动状态。

1.3.1 加速度计

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。

传统加速度计的工作原理主要基于质量和弹簧的相互作用。当物体受到加速度时，内部的质量块会受到相应的力而发生位移，这个位移会导致弹簧或其他弹性元件发生变形。通过测量这种变形量，可以推算出物体所受的加速度大小。

现代加速度计，如 MEMS（微机电系统）加速度计，则采用了不同的工作原理。MEMS 加速度计通常包含三个主要部分：上电容板、中电容板（可移动）和下电容板。当有加速度时，中电容板会根据加速度的方向和大小发生移动，从而改变与上、下电容板之间的距离，进而改变电容的容值。这种电容值的变化与加速度成正比，通过测量电容

值的变化并经过数字处理，可以输出数字化的加速度信号。

加速度计的应用场景非常广泛。在飞机、导弹等航空器中，加速度计被用于姿态控制，帮助飞行员或自动驾驶系统实时了解飞行器的姿态信息，以便进行精确的操纵。在卫星、载人飞船等航天器中，用于轨道控制等任务。同时，在汽车行业中，加速度计被广泛应用于车辆稳定性控制、碰撞检测等领域。通过测量车辆的加速度变化，可以实时评估车辆的行驶状态，并在必要时采取相应的控制措施，以提高车辆的安全性和稳定性。并且，在智能手机、平板电脑等消费电子产品中，加速度计也扮演着重要的角色。用于实现横竖屏切换、计步等功能。这些功能不仅提高了用户的体验，还使得产品更加智能化和便捷化。

1.3.2 陀螺仪

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。

陀螺仪作为一种能够维持其旋转轴稳定定向的设备，其工作原理基于角动量守恒定律。陀螺仪内部通常包含一个高速旋转的转子，当转子受到外力作用时，其旋转轴会保持相对稳定的方向不变。这种稳定性使得陀螺仪能够在各种复杂环境中提供精确的方向信息。

陀螺仪的应用场景同样非常广泛，在飞机和航天器中，陀螺仪被用于测量和维持方向。飞行员和自动驾驶系统可以依靠陀螺仪提供的姿态信息来进行精确的操纵和导航。同时，在导弹等武器系统中，陀螺仪也扮演着重要的角色，用于确保导弹在飞行过程中的稳定性和准确性。在船舶和潜艇中，陀螺仪被用于确定航向和姿态。通过测量船舶或潜艇的角速度变化，陀螺仪可以提供精确的方向信息，帮助船员进行导航和定位。同时，由于惯性导航系统是一种不依赖于外部信号源（如 GPS）的导航系统。它利用陀螺仪和加速度计等传感器来测量物体的运动状态（包括角速度和加速度），并通过积分运算计算出物体的位置。这种系统在 GPS 信号无法接收的环境中（如地下、室内、深海或深空）具有特别重要的意义。并且，在智能手机、游戏控制器等消费电子产品中，陀螺仪也被广泛应用。例如，在游戏控制器中，陀螺仪可以感应玩家的手部动作并转化为游戏角色的运动指令；在智能手机中，陀螺仪则可以被用于实现屏幕自动旋转等功能。

1.4 本文研究内容和组织结构

本文主要分为 5 个部分。第一部分是绪论，介绍了研究的相关背景和现状，以及现有的行为识别所存在的不足，并进一步引出了本文要做的改进工作。第二部分简单介绍了本文使用的数据预处理方法及选用的特征值。第三部分介绍了所使用的算法基本内容与其优缺点。第四部分介绍了结果分析及前景展望。

接下来，解决以下三个具体问题

1.根据附件 1 中 3 名实验人员的运动数据，利用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，对每一位实验人员的活动状态的数据进行分类，得出分类结果。

2.根据附件 2 中 10 名实验人员的活动数据，利用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，提取了 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，并利用模型开展了以下验证工作：

（1）进一步运用问题 1 的分类模型对该 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），比较了问题 2 中判别模型和问题 1 的分类模型的结果，分析了采用分类模型对不同活动类型分类时的分类准确度。

（2）利用附件 3 中收集到的某实验人员 30 次活动的状态数据，运用判别模型，给出了该人员的活动状态。

3.根据附件 4 给出的问题 1 和问题 2 中参与实验的 13 位实验人员的年龄、身高、体重等数据，分析了不同人员的同一活动状态存在差异。活动状态数据与实验人员的年龄、身高、体重有关系，能使用活动传感器数据进行人员画像。并且，进一步根据附件 5 中给出的问题 2 的 10 位实验人员中的 5 位的某次活动数据，利用模型判断他们分别最可能来源于问题 2 中对应的实验人员。

最后，总结了本文所做的工作，提出了下一步研究展望。

2 数据处理与特征选择

2.1 数据预处理

需要指出的是，本文采用笛卡尔坐标系，定义 x 轴正方向为重力方向，y 轴正方向为人体前进方向，z 轴正方向为人体右侧方向。

对于一般的惯性导航应用场景，轨迹解算依赖于地球坐标系。在已知起始状态时根据惯性传感器件的参数，进行姿态分析与轨迹判断。但是本文提及的场景，无法获取人体的起始状态，因而一般的惯性导航算法无法直接使用。本文选取了一些不依赖初始姿态的特征值，例如合加速度、转动角度等，这些数据有利于增强所建模型的物理意义及可解释性，在模型求解中起到了至关重要的作用。值得注意的是，题目中给出的一些运动场景，如向左转、向右转，步行上、下楼等，可能依赖于人体的初始姿态，所以在实际的模型判别中会遇到困难。

2.1.1 合加速度

加速度计通常能够测量物体在三个互相垂直的轴（如 x 轴、y 轴、z 轴）上的加速度分量。这三个加速度分量分别代表了物体在这三个方向上的速度变化率。

合加速度（也称为总加速度或矢量和加速度）是这三个加速度分量在三维空间中的矢量和。它表示了物体在任意方向上速度变化的总体效果。在数学上，合加速度可以通过矢量加法的方式来计算，即使用勾股定理（或其三维空间中的推广）来计算。

具体地，如果 a_x、a_y、a_z 分别表示物体在 x 轴、y 轴、z 轴上的加速度分量，那么合加速度 a 的大小可以通过以下公式计算：

$$a = \sqrt{a_x + a_y + a_z} \tag{1}$$

合加速度表示了物体在任意方向上速度变化的快慢和方向。它是描述物体运动状态变化的重要物理量之一。通过测量和分析合加速度，可以了解物体在三维空间中的运动特性，如速度的变化趋势、加速度的方向等。

2.1.2 低通滤波

低通滤波去噪方法的主要流程如图 4 所示。在实际应用中，惯性传感元件易受到噪声扰动，容易造成姿态的误分析。结合人体运动特点，一般的运动类型，其主要动作频率不会高于 3Hz，对测量信号进行低通滤波，可以减小噪声对运动类型识别的干扰。主要步骤如下:①对带有噪声的信号进行傅里叶变化处理，以获得频域频域信号。②选择无限冲激响应低通滤波器。③选择合适的截止频率和阶数将噪声产生的高频信号删除，本文采样频率为 100Hz，截至频率为 3Hz。④进行逆傅里叶变换，以获得去噪后的信号。

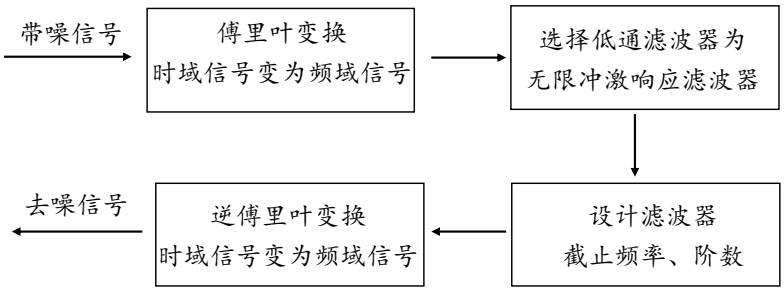


图 4 低通滤波去噪的流程图

对合加速度数据和陀螺仪数据进行低通滤波处理，得到去除噪声后的信号。并对加速度计数据进行计算，得到合加速度。

2.1.3 异常值与缺失值

利用多重集值信息系统处理不完备数据集的异常点检测问题[10]。首先，将不完备数据集转换为多重集值信息系统。然后，基于异常点检测模型，利用信息粒化，定义了多重集值信息系统中的异常点，并采用了相应的算法对异常值进行剔除，然后插值补充了缺失值[11]。

2.1.4 数据的归一化与标准化

为进行不同特征值的分析，防止不同类特征值的数值量级差异过大带来分析误差，对计算得到的特征数据进行了归一化操作。

对每类特征值归一化操作具体流程为，①获取其最大正值，以之为 1。②对于最小负值，以之为-1。③以此作为基准，将特征值的分布线性变换至[-1，1]区间。

为进一步研究不同人之间的活动情况，进行数据的标准化过程是必要的。本文选择了标准分数（Z-score）标准化方法，对每个特征计算其平均值与标准差，将每个特征的值减去均值再除以标准差。Z-score 方法可以降低不同类型特征之间的量纲和变化范围带来的影响，增强数据的可比性。通过归一化和标准化，有利于减小数据偏差，提高数据的质量，保证建模的精确度。

2.1.5 主成分分析（PCA）

主成分分为主成分分析和主成分评价两个方面，分析是单纯的分析数据是否具有主成分和主成分效果如何，评价是根据主成分运行的结果直接评价。当指标过多，并且各个指标之间相互有影响，为了消除指标之间的影响，单纯从数据的角度寻找各个指标具有公共特征，这些公共的特征就是主成分，也就是常说的第一主成分。

PCA 沿着与样本观察值具有最大数量变异的方向对数据进行简化，从而将数据的描述功能的复杂性降低到最低程度。在空间上，PCA 可以理解为将原始数据投射到一个新的坐标系统，其中第一主成分对应第一坐标轴，表示了原始数据中多个变量通过某种变换得到的新变量的变化范围。通过这种方式，PCA 能够提取数据中的主要变化模式，并忽略掉次要的变化，从而实现数据的降维。在数据可视化中可以更好地进行数据展示和分析，在机器学习中可以更好地进行模型训练和预测。

虽然 PCA 会损失一些次要的信息，得出的主元可能并不是最优的。但是 PCA 通过完全无参数限制地提取主成分，能够去除数据中的冗余信息，保留最重要的特征，实现数据降维，这显著提高了计算效率。

具体流程为：①数据收集与预处理：收集原始数据，并进行必要的清洗和预处理，去除异常值、缺失值填充等。②数据标准化：对预处理后的数据进行标准化处理，以确保不同特征具有相似的尺度。③计算协方差矩阵：计算标准化后数据的协方差矩阵。④求解特征值和特征向量：对协方差矩阵进行特征值分解，得到特征值和对应的特征向量。⑤选择主成分：根据特征值的大小，选择前几个特征向量作为主成分。⑥构建新的特征空间：使用选定的主成分构建新的特征空间。

2.2 特征提取与特征选择

为了提高预测的准确性，利用特征提取与特征选择来构建更快、消耗更低的预测模型。特征提取和特征选择通过选择重要特征的子集来减少数据的维度，以增强模型的可解释性[12]。对于本应用场景，依据数据的物理意义选取了几种重要的特征，如用以表征运动情况的合加速度平均分段方差，用以衡量左右转动的角度积分等。

2.2.1 合加速度平均分段方差

因为有重力加速度的存在，所以通过加速度计测得的数据无法确定加速度初始值。因此选取合加速度的方差这一参考变量来判断是否处于运动状态。

结合人体的运动规律，为检测出某时间段的运动与否，可以采用 200 个数据大小的窗口进行分析。当人体处于静止时，该窗口内的合加速度方差很小；随着运动的持续，该窗口内的合加速度方差逐渐增大。合加速度的方差不仅可以反应出人体的运动与否，还可以反应出运动的剧烈强度，从而有助于静止、走路、跑动的分类。合加速度分析的主要流程如图 5 所示。

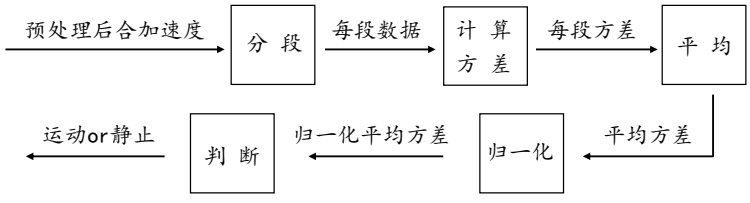


图 5 合加速度分析流程图

主要步骤如下:①对去噪后数据进行分段，每 200 个数据即 2s 为一段。②分别计算每段的方差。③计算方差的平均值。④对平均方差归一化。⑤当归一化平均方差小于 0.1 时，认为总体状态为相对静止状态。当归一化平均方差为 0.1-0.5 之间时，认为总

体状态为平稳运动状态。当归一化平均方差为 0.5-0.8 时，认为总体状态为剧烈运动。当归一化平均方差大于 0.8 时，认为总体状态为超剧烈运动。其中，运动和静止的具体活动分类见表 1。

表 1 运动和静止活动分类

运动状态	相对静止	平稳运动	剧烈运动	超剧烈运动
具体活动	乘坐电梯向上移动、乘坐电梯向下移动、坐下、站立、躺下	向前走、向左走、向右走、步行上楼、步行下楼	向前跑	跳跃
判断标准	＜0.1	0.1-0.5	0.5-0.8	＞0.8

2.2.2 旋转角度

因为旋转轴可能随时间变化，对角速度进行积分得到的是旋转向量的角度部分，而不是完整的旋转向量。

陀螺仪所记录角速度分析的主要流程如图 6 所示。主要步骤如下:①使用数值积分方法对角速度进行积分，以得到每个时间步长的旋转向量增量。假设在短时间内旋转轴不变或变化很小。因此，可以将角速度向量视为旋转轴，并将其模长（即角速度的大小）乘以时间步长来近似得到旋转角度的增量。②将每个时间步长的旋转向量增量转换为四元数增量。③使用四元数乘法来更新元数表示的姿态。④使用四元数来表示和累加旋转，以得到总旋转角度。

通过计算绕 x 轴的旋转角度，可以实现运动方向的区分。

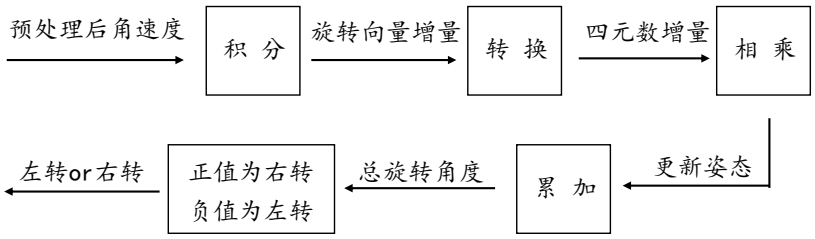


图 6 角速度分析流程图

2.2.3 其他特征

为补充特征值种类，增进分类效果，经过广泛调研及分析后又另选取了如下 30 种特征量，如表 2 所示。

表 2 时域特征、频域特征提取

数据	时域特征	频域特征
各列测量值、合加速度	最小值、最大值、均值、众数、四分位数(3 个)、四分位距、极差、方差、平均绝对离差、能量	最小值、最大值、均值、四分位数(3 个)、四分位距、极差、方差、标准差、均方根、平均绝对离差、能量
加速度计 三轴之间	皮尔逊相关系数 (3 个值)	无
陀螺仪 三轴之间	皮尔逊相关系数 (3 个值)	无

3 算法实现

在分类算法设计上，近年来机器学习算法得到了广泛应用并取得良好效果，如朴素贝叶斯、支持向量机、K 近邻、决策树、随机森林、神经网络等。通常，流量分类场景中特征向量复杂度远少于图像处理、自然语言处理等场景，因此一般不必采用太复杂的机器学习算法（或深度学习）即可取得良好效果。

3.1 K-means 分类

K 近邻[13][14] (K-means)常用于进行无监督分类。通过距离来衡量数据样本间的相似程度。负荷曲线样本间的距离越小，负荷曲线越相似，在同一簇类的可能性越大。计算数据样本间距离的方法有很多种，K-means 算法通常采用欧氏距离来计算数据样本之间的距离。这种物理表示方式使得算法能够直观地反映数据点之间的相似性和差异性，从而实现有效的聚类。

具体流程为：①初始化：随机选择 K 个点作为初始聚类中心。②分配：对于数据集中的每个点，找到最近的聚类中心，并将该点分配给对应的聚类。③更新聚类中心：根据聚类中所有点的均值，重新计算每个聚类的中心。④重复步骤 2 和 3，直到聚类中心不再显著变化或达到预设的迭代次数。

3.2 随机森林

随机森林(Random Forest，RF)是一种使用多棵决策树进行训练和预测的分类器，从原始的训练数据中有放回地抽取 k 个样本构成m个训练集，在每个训练集上构造一棵决策树。构造决策树时，在节点寻找特征进行分裂时，随机选取一些特征，在抽取到的特征中选择最优解进行分裂，最终根据多数投票原则产生分类结果。

具体流程为：①随机森林参数设置：首先，确定决策树数量。通常数量越多，模型的性能会越好，但同时也会增加计算成本。然后，选择特征子集。对于每个决策树，从原始特征集合中随机选择特定数量的特征子集进行构建。以增加模型的多样性，减少各个决策树之间的相关性。接下来，选择样本子集。从原始样本集合中随机选择样本子集，用于构建每个决策树。本文使用有放回的随机抽样方式，以增加模型的多样性。②构建决策树：使用随机选取的特征子集和样本子集，为每个决策树构建过程提供输

入。使用基于信息增益、基尼系数或其他衡量指标的算法来构建决策树。③集成决策树：将所有决策树的预测结果进行集成。对于分类问题，通常使用投票机制，即根据每个决策树的分类结果进行投票，选取得票最多的类别作为最终的分类结果。④模型评估与调优：使用评估指标对随机森林模型进行评估。根据评估结果对模型进行调优。尝试调整决策树的数量、特征子集的大小、样本子集的大小等参数，以提高模型的性能和预测准确度。

3.3 支持向量机

支持向量机（Supporting Vector Machine，SVM）[15]作为一种有监督学习算法，主要用于分类问题[16]。具有高维数据有效、泛化能力强、鲁棒性好和可解释性强等优点。其核心思想是在特征空间中寻找一个间隔最大的分离超平面，以最大化不同类别数据之间的间隔，从而实现分类。

不失一般性，假设大小为 l 的训练样本集{（x_i，y_i），i=1，2，……，l}由两个类别组成，若 x_i 属于第一类，则记 y_i=1；若 x_i 属于第二类，则记 y_i=-1。

若存在分类超表面

$$wx+b=0 \tag{2}$$

能够将样本正确地划分成两类，即相同类别的样本都落在分类超平面的同一侧，则称该样本集是线性可分的。即满足

$$\begin{cases} wx_i+b\geq 1, & y_i=1 \\ wx_i+b\leq -1, & y_i=-1 \end{cases}, \quad i=1,2,\cdots,l \tag{3}$$

定义样本点 x_i 到式 3 所指的分类超平面的间隔为

$$\varepsilon_i=y_i(wx_i+b)=\left|wx_i+b\right| \tag{4}$$

将式(28-3)中的 w 和 b 进行归一化，即用 w/‖w‖和 b/‖w‖分别代替原来的 w 和 b，并将归一化后的间隔定义为几何间隔

$$\delta_i=\frac{wx_i+b}{\|w\|} \tag{5}$$

同时，定义一个样本集到分类超平面的距离为此集合中与分类超平面最近的样本点的几何间隔，即

$$\delta=\min\delta_i, \quad i=1,2,\cdots,l \tag{6}$$

样本的误分次数 N 与样本集到分类超平面的距离 δ 间的关系为

$$N\leq\left(\frac{2R}{\delta}\right)^2 \tag{7}$$

其中，R=max‖x_i‖，i=1，2，……，l，为样本集中向量长度最长的值。

由式 8 可知，误分次数 N 的上界由样本集到分类超平面的距离 δ 决定，即 δ 越大，N 越小。因此，需要在满足式 4 的无数个分类超平面中选择一个最优分类面，使得样本集到分类超平面的距离 δ 最大。

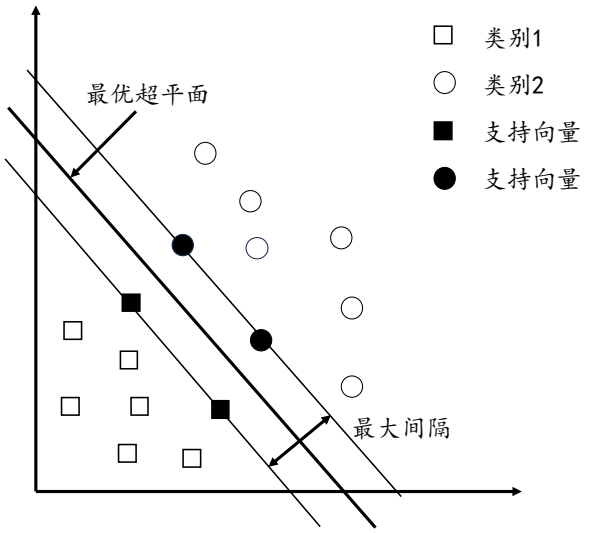


图 7 最优超平面示意图

若间隔 $\varepsilon=|wx_i+b|=1$，则两类样本点间的距离为 $2|wx_i+b|/\|w\|=2/\|w\|$。因此，如图 2.3 所示，目标即为在满足式 4 的约束下寻求最优分类超平面，使得 $2/\|w\|$最大，即最小化$\|w\|^2/2$。用数学语言描述，即

$$\begin{cases} \min \dfrac{\|w\|^2}{2} \\ s.t. y_i(wx_i+b) \geq 1, \quad i=1,2,\cdots,l \end{cases} \tag{8}$$

该问题可以通过求解 Largrange 函数的鞍点得到，即

$$\phi(w,b,a_i)=\frac{1}{2}\|w\|^2-\sum_{i=1}^{i} a_i[y_i(wx_i+b)-1] \tag{9}$$

其中，$a_i>0$，$i=1$，2，…，l，为 Largrange 系数。

由于计算的复杂性，一般不直接求解，而是依据 Largrange 对偶理论将式 10 转化为对偶问题，即

$$\begin{cases} \max Q(a)=\sum_{i=1}^{l} a_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} a_i a_j y_i y_j (x_i x_j) \\ s.t. \sum_{i=1}^{l} a_i y_i = 0, \quad a_i \geq 0 \end{cases} \tag{10}$$

这个问题可以用二次规划方法求解，设求解得到的最优解为 $a^*=[a_1^*, a_2^*, \cdots, a_l^*]^T$，则可以得到最优的 w^*和 b^*为

$$\begin{cases} w^* = \sum_{i=1}^{l} a_i^* x_i y_i \\ b^* = -\frac{1}{2} w^* (x_\alpha + x_\beta) \end{cases} \tag{11}$$

其中，x_α和 x_β为两个类别中任意的一对支持向量。

最终得到的最优分类函数是

$$f(x) = \mathrm{sgn}[\sum_{i=1}^{l} a_i^* y_i (xx_i) + b^*] \tag{12}$$

值得一提的是，若数据集中的绝大多数样本是线性可分的，仅有少数几个样本(可能是异常点)导致寻找不到最优分类超平面。针对此类情况，通用的做法是引人松弛变量，并对式 9 中的优化目标及约束项进行修正，即

$$\begin{cases} \min \frac{\|w\|^2}{2} + C\sum_{i=1}^{l} \xi_i \\ s.t. \begin{cases} y_i(wx_i + b) \geq 1 - \xi_i \\ \xi_i > 0 \end{cases}, \ i = 1, 2, \cdots, l \end{cases} \tag{13}$$

其中，C 为惩罚因子，起着控制错分样本惩罚程度的作用，从而实现在错分样本的比例与算法复杂度间的折中。求解方法与式 10 相同，即转化为其对偶问题，只是约束条件变为

$$\begin{cases} \sum_{i=1}^{l} a_i y_i = 0 \\ 0 \leq a_i \leq C \end{cases}, \ i = 1, 2, \cdots, l \tag{14}$$

最终求得的分类函数的形式与式 13 一致。

3.4 朴素贝叶斯

朴素贝叶斯（Naive Bayes， NB）作为基于贝叶斯定理和特征条件独立假设的分类方法，使用贝叶斯公式计算后验概率，并选择后验概率最大的类别作为预测结果。每个特征独立地对分类结果产生影响，通过计算各特征的后验概率来进行分类。虽然特征条件独立假设在现实中往往不成立，可能导致分类性能下降。但是具有计算速度快，模型简单，易于实现等优点。

具体流程为：①训练分类器：为每一个类别计算其先验概率，即样本属于每个类别的概率。并针对每个类别，计算在该类别下各特征出现的概率。这里假设特征之间

相互独立，因此可以将联合概率分解为各个特征概率的乘积。根据贝叶斯定理和前面计算的概率，构建朴素贝叶斯分类器。分类器的输入是待分类项的特征向量，输出是待分类项与类别的映射关系。②应用分类器：将待分类项的特征向量输入到训练好的朴素贝叶斯分类器中。并计算待分类项属于每个类别的后验概率。最终选择后验概率最大的类别作为分类结果。③评估与优化：使用测试集评估分类器的性能。根据评估结果调整特征选择、参数设置等，以优化分类器的性能。

3.5 决策树

决策树(Decision Tree， DT)通过递归地选择最佳特征来划分数据集，构建树形结构进行分类。通常使用信息增益、基尼指数等作为划分标准。通过树形结构表示决策过程，每个节点代表一个特征，每个分支代表该特征的不同取值，叶子节点代表分类结果。虽然该方法具有模型直观、易于理解、分类速度快等优点，但是容易过拟合、对输入数据的微小变化可能过于敏感。

具体流程为：①决策树生成：首先，从所有特征中选择一个最优特征作为根节点，这个特征应该具有最高的信息增益。然后，根据根节点的特征值将数据集分割成不同的子集，每个子集对应树的一个分支。接下来，对每个分支重复上述步骤直到满足停止条件。停止条件设定为节点的样本数小于预设阈值。②剪枝：在决策树生成过程中，先进行预剪枝，通过设定节点最小样本数来限制树的复杂度，提前停止树的生长。在决策树完全生成后，进行后剪枝，基于损失函数优化实现。通过删除一些子树或叶节点来简化树的结构，降低过拟合的风险。③模型评估：使用训练集数据对生成的决策树进行初步评估，了解模型在已知数据上的表现。使用独立的测试集数据对模型进行评估，以检验模型的泛化能力。其中，评估指标为准确率、召回率和 F1 分数。④应用与调整：将训练好的决策树模型应用于实际问题中，进行预测或分类。根据模型在测试集上的表现，调整最大深度、最小样本数等模型参数，以进一步优化模型性能。

3.6 判别分析

判别分析（Discriminant Analysis， DA）是一种监督学习算法，用于发现数据集中不同群组之间的差异，并利用这些差异对新观测值进行分类。线性判别分析（Linear Discriminant Analysis， LDA）和二次判别分析（Quadratic Discriminant Analysis， QDA）是两种主要的判别分析方法。其中线性判别分析假设不同类别的数据具有相同的协方差矩阵，判别函数是自变量的线性组合；二次判别分析，允许每个类别的数据具有不同的协方差矩阵，判别函数是自变量的二次函数。

具体流程为：①建立判别模型：选择适当的判别分析方法（LDA 或 QDA）；计算每个类别的均值和协方差矩阵；确定判别函数的系数，构建判别模型。②模型评估：使用训练数据集评估模型的拟合优度；通过交叉验证等方法评估模型的预测准确性。③模型诊断：检查模型的假设条件是否满足，如多元正态分布、方差齐性等；识别并处理潜在的多重共线性问题。④模型优化：根据模型评估和诊断的结果，对模型进行调整和优化；可能包括变量选择、模型参数调整等。

其存在一定的缺点，比如，判别分析要求数据在每个类别中都服从多元正态分布，这在实际应用中可能不总是成立。另外，线性判别分析要求所有类别具有相同的协方差矩阵，这限制了其在方差不齐性数据上的使用。但是其优点是很明显的：方法直观，

判别分析结果易于解释，特别是当使用线性判别分析时，判别函数的系数可以直观地显示哪些自变量对分类结果有显著影响；预测能力，判别分析可以很好地预测新数据的类别，尤其是在数据满足其假设条件时。

4 结果分析

4.1 问题 1

4.1.1 分类模型及其分析

首先选用常见的 30 个特征值，利用 K-means 进行聚类。该方法无需数据标签，适用于解决问题 1。对特征进行了主成分分析（PCA）后，选用 10 个主成分，将数据进行聚类。图 8 示出第一、二主成分的散点图。

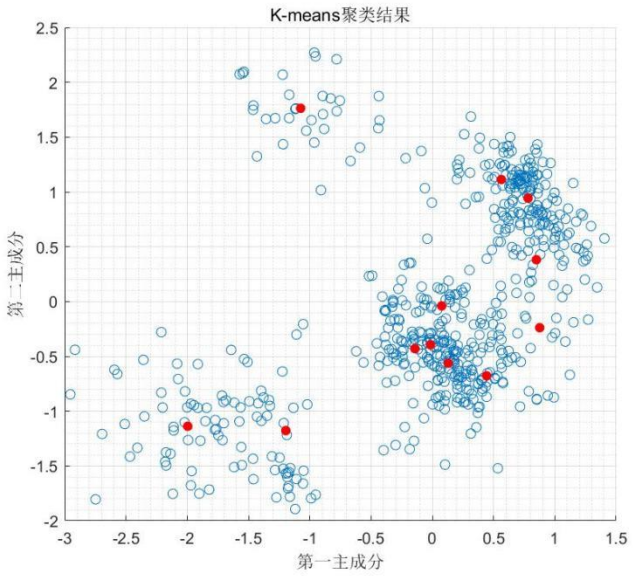


图 8 K-means 聚类，红色为聚类中心。

4..1.2 分类结果

实验结果表明，仅使用这些特征值难以进行较好地分类。结合我们设计的“合加速度平均方差”，“转动角度”两个特征量，我们获取了如下的分类结果，如表 3 所示。

从表 3 中可以看到，加入我们设计的特征之后，分类结果较为均匀。结果表明，第二类、三类、四类和五类，以及第七类、八类和十一类容易出现分类偏差。值得指出的是，K-means 方法依赖于初值的选择，其分类存在一定的随机性。

表 3 问题 1 结果

分类	Person1	Person2	Person3
第一类	SY1, SY2, SY15, SY23, SY36	SY1, SY4, SY48, SY49, SY50	SY1, SY7, SY28, SY40, SY45
第二类	SY3, SY9, SY10, SY18, SY57, SY60	SY2, SY10, SY12, SY26, SY47	SY2, SY6, SY12, SY33, SY34, SY47
第三类	SY4, SY13, SY24, SY39	SY3, SY20, SY38, SY51, SY60	SY3, SY32, SY50, SY54, SY55
第四类	SY5, SY8, SY56, SY59	SY5, SY11, SY22, SY54, SY56	SY4, SY5, SY53
第五类	SY6, SY12, SY45, SY51, SY55	SY6, SY15, SY25, SY35, SY45, SY53, SY58	SY8, SY23, SY39, SY42, SY52
第六类	SY7, SY33, SY38, SY42, SY58	SY7, SY30, SY33, SY44, SY57	SY9, SY14, SY30, SY37, SY43
第七类	SY11, SY16, SY20, SY30, SY54	SY8, SY14, SY18, SY55	SY10, SY18, SY22, SY41, SY57, SY58, SY59
第八类	SY14, SY27, SY31, SY37, SY49	SY9, SY17, SY36, SY41	SY11, SY25, SY31, SY44
第九类	SY17, SY28, SY41, SY47, SY52	SY13, SY27, SY28, SY34, SY42	SY13, SY19, SY24, SY29, SY49
第十类	SY19, SY26, SY43, SY44, SY50	SY16, SY21, SY29, SY40, SY43	SY15, SY16, SY20, SY26, SY46
第十一类	SY21, SY25, SY29, SY46, SY48, SY53	SY19, SY23, SY24, SY37, SY46	SY17, SY21, SY38, SY48
第十二类	SY22, SY32, SY34, SY35, SY40	SY31, SY32, SY39, SY52, SY59	SY27, SY35, SY36, SY51, SY56, SY60

4.2 问题 2

4.2.1 判别模型

选用了 matlab 中常用的分类器模型 KNN、Kernel、Linear、NB、DT、SVM 和 DA 对附件 2 中的 9 组数据训练判别模型，并选用其余一组检验判别模型的准确性。为增加结果的普遍性，进行了多次实验，每次实验随机选取 10 组数据中的某一组作为测试组。检验结果表明，Linear 方法可以表现出较好的准确性。

表 4 示出上述不同方法的准确性。可以看到，使用 Linear 方法时，可以获得更高的准确率。后文的判别模型是基于该方法进行构建的。

表 4 不同方法的判别准确率

核类型	准确率	核类型	准确率
NB	83.33%	Linear	93.33%
DT	82.53%	SVM	91.67%
DA	91.67%	Kernel	50.0%
KNN	90.83%		

4.2.2 分类模型对不同活动类型的分类准确度

利用问题 1 分类模型对附件 2 中 10 名实验人员数据进行分类，结果如图 9 所示。

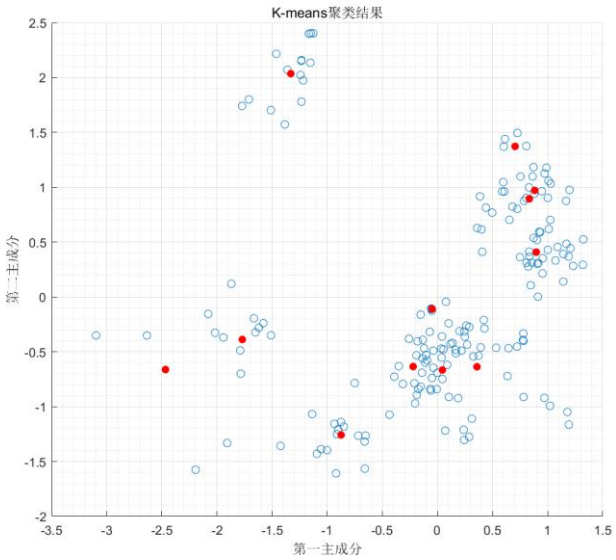


图 9 问题 1 模型分类结果

问题 1 中的分类模型与问题 2 中判别模型所得分类结果的符合度为 74.33%。针对 12 种运动类型，分类模型与判别模型预测的具体结果如图 10 所示。

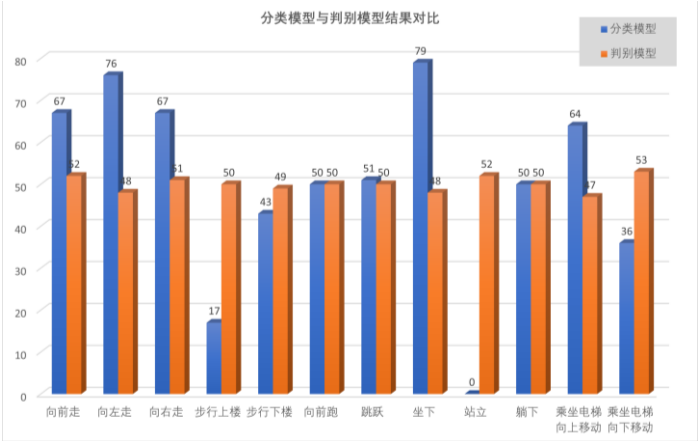


图 10 分类模型与判别模型效果对比

经过分析，问题 1 选择的模型对于“躺下”、“跳跃”、“跑步”的划分是最为精准的，这得益于本文选择的“合加速度平均方差”特征。另外可以看到，其他相对静止（如乘坐电梯、坐下、站立）和平稳运动（向前走、向左走、向右走以及步行上下楼）直接的划分也是清楚的。由此可以看出，本文构建的“合加速度平均方差”特征可以将运动强度清楚地划分为“相对静止”、“平稳运动”、“剧烈运动”、“超剧烈运动”。经过试验，单独使用这一特征量可以在很大程度上将活动状态进行以上四类划分，具有很重要的应用价值。

经过判断，分类的误差多数出现在上、下楼，左、右转，坐下、站立和坐电梯上、下之间。这些数据的特点是上下楼，左右转之间，其运动的强度基本一致，合加速度平均方差应当处于相近的量级。用来区分左、右转的旋转角度也很容易受到测量噪声的影响，即使使用了低通滤波后，效果仍然不尽如人意。坐电梯上、下楼之间，可以通过超重、失重出现的时机进行判别。但是实际的电梯运行总有振荡过程，导致这种判别特征容易被湮没。坐下、站立之间，也存在相近的问题。

4.2.3 判别模型结果

本文利用训练好的判别模型，对附件 3 中的 30 次实验活动数据进行了判别，结果如表 5 所示。

表 5 问题 2 结果

活动类型	判别状态	活动类型	判别状态	活动类型	判别状态
SY1	2	SY11	9	SY21	6
SY2	1	SY12	7	SY22	2
SY3	7	SY13	4	SY23	12
SY4	11	SY14	3	SY24	5
SY5	7	SY15	4	SY25	2
SY6	10	SY16	1	SY26	9
SY7	2	SY17	4	SY27	8
SY8	6	SY18	5	SY28	5
SY9	7	SY19	8	SY29	6
SY10	10	SY20	8	SY30	2

4.3 问题 3

问题 2 中构建的判别模型输出结果为运动的类型，因此在解决问题 3 时，模型需要进行调整以使得输出结果为人员的编号。

对于不同人员，其各自运动的特点有其特征。对于进行同一运动的不同个人来说，合加速度最大值、方差值等特征值表现出了差异性，这为判别带来了可能。在选取的运动特征基础上，本文又分别向判别模型中添加了年龄、身高、体重特征及其组合作为补充，验证这些特征与运动判别的关联。利用附件 2 中的 10 位实验人员中的 5 位作为训练集，剩余 5 位作为验证集。模型的准确率如图 11 所示。

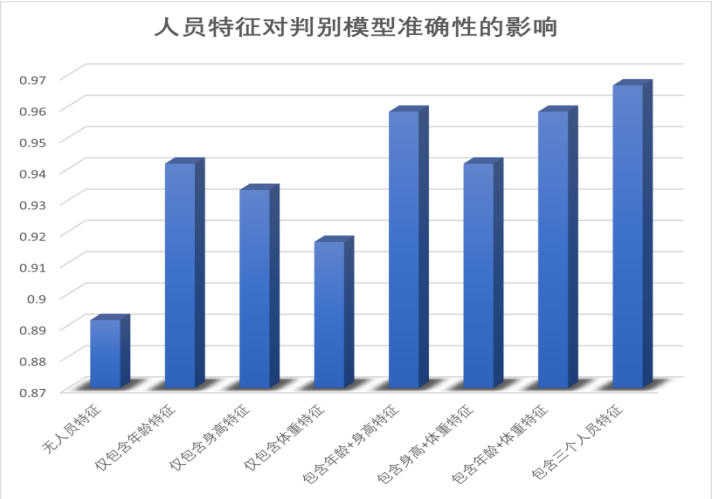


图 10 人员特征对判别模型准确性的影响

实验结果表明，①在未引入人员年龄、身高及体重信息时，判别模型的准确性最低。②在单独添加了年龄、身高和体重后，模型的判别准确性有所提升。仅引入体重信息对于模型判别的提升量最小。③在将人员特征两两结合后，模型的准确率进一步提高。④引入所有的人员特征后，模型判别准确率达到最高。

这证明了判别模型的准确性是与人员特征密切相关。一方面，借助人员特征信息，可以很好地提升运动状态判别的准确性；另一方面，在可以掌握大量数据的情况下，可以在很大程度上训练判别模型，通过运动特征进行人员画像。

为进一步地从五位实验人员的活动数据中判断出他们的具体编号，预测方法如下：①首先选取问题 2 中提及的 10 个实验人员，将十二种运动状态分别进行有监督的训练，得到 12 个分类器。该分类器的输出为人员编号。②利用这 12 个分类器分别对未知人员的 12 种运动状态进行预测，得到 12 个人员编号预测结果。③选择这 12 个人员编号预测结果中出现最多的人员编号作为本次对未知人员的预测结果。

表 6 问题 3 结果

活动类型	判别结果
Unkonw1	10
Unkonw2	7
Unkonw3	6
Unkonw4	9
Unkonw5	13

5 展望

基于加速度计和陀螺仪数据的活动状态识别技术，正逐步成为推动健康生活、智能家居、工业安全、专业体育训练及娱乐游戏等多个领域创新发展的关键力量。该技术通过高精度地捕捉并分析人体运动数据，能够实时识别用户的活动模式，为个性化服务、安全监控及效率提升提供了强有力的数据支持。

展望未来，这项技术将进一步融合深度学习算法，提升对复杂活动模式的识别精度与泛化能力。同时，多源数据融合技术的发展将使得加速度计和陀螺仪数据能够与其他传感器数据（如 GPS、心率监测等）无缝结合，构建出更加全面、精准的用户活动画像。此外，随着隐私保护意识的增强，该技术也将更加注重数据加密与匿名化处理，确保用户数据的安全与隐私。在跨平台标准化方面，统一的 API 接口和数据标准将促进技术的普及与应用，进一步拓展其市场潜力和影响力。

参考文献

[1] Aras Y， Billur B. Novel Noniterative Orientation Estimation for Wearable Motion Sensor Units Acquiring Accelerometer， Gyroscope， and Magnetometer Measurements [J]. IEEE Transactions on Instrumentation & Measurement， 2020， 69(6): 3206-3215.

[2] Chen P， Li Z， Zhang S.Indoor PDR Trajectory Matching by Gyroscope and Accelerometer Signal Sequence without Initial Motion State [J]. IEEE Sensors Journal， 2023，23(13): 1.

[3] Ma H， Li W， Zhang X， et al. AttnSense: Multi-level Attention Mechanism For Multimodal Human Activity Recognition[C].IJCAI，2019: 3109-3115.

[4] Nweke H F， The Y W， Mujtaba G， et al. Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions [J]. Information Fusion， 2019， 46: 147-170.

[5] 田佳婕，张雨欣，肖毅.我国可穿戴体育设备发展现状及趋势分析[A].第十三届全国体育科学大会[C]，2023-11-03.

[6] 孟政元.iSense：基于智能手机传感器的行为识别应用研究[D].吉林大学，2022.

[7] 张富春，郑武略，汪豪，等.基于改进 EKF 的柔性绳索自动升降装置姿态估计[J].兵器装备工程学报，2020，41(11): 230-236.

[8] Li K， Habre R， Deng H， et al. Applying Multivariate Segmentation Methods to Human Activity Recognition From Wearable Sensors' Data. [J]. JMIR mHealth and uHealth，2019， 7(2): e11201.

[9] Hsu Y C， Wang H， Zhao Y， et al. Automatic Recognition and Analysis of Balance Activity in Community-Dwelling Older Adults: Algorithm Validation [J]. Journal of Medical Internet Research，2021， 23(12): e30135.

[10] 林海.多重集值信息系统及其在缺失数据处理中的应用[D].广西大学，2023.

[11] 邹同华，高云鹏，伊慧娟，等.基于 Thompson tau-四分位和多点插值的风电功率异常数据处理[J].电力系统自动化，2020， 44(15): 156-165.

[12] 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.

[13] Cai Y D， Huang J Z， Yin J F. A new method to build the adaptive k-nearest neighbors similarity graph matrix for spectral clustering[J].Neurocomputing， 2022， 493: 191-203.

[14] Prasad R K， Sarmah R， Chakraborty S， et al. Sauravjyoti Sarmah.NNVDC: A new versatile density-based clustering method using k-Nearest Neighbors[J]. Expert Systems with Applications，2023， 227: 120250.

[15] Zhang J Q， Yang H.Bounded quantile loss for robust support vector machines-based classification and regression[J].Expert Systems with Applications，2024，Vol.242: 122759

[16] 宋慧玲，张仲广，夏冰著.基于支持向量机的分类问题研究[M].黑龙江大学出版社，2022.

附　录

附录 A：问题 1 和问题 2 代码

首先利用 get_date 程序完成问题 1、2 获取数据部分，然后建立模型 build_mode1，最终检验模型 text_mode1。

get_date.m 代码

get_data.m

```
%%问题1 用来读取数据，完成数据预处理和获取数据特征矩阵data
clear
paths;
glvs;
A = zeros(32,60);B = zeros(32,60);
frist_name = ['test1';'test2';'test3';'test4';'test5'];
fs = 100; % 采样率为100 Hz
ft = 200; % 数据分割频率

% figure
flag = zeros(60,3);label = [];data = [];
for nn = 4:13
for i1 = 1:12
    for i2 = 1:5
    i=(i1-1)*5+i2;
%% 数据初处理
F = xlsread([frist_name(2,:),'\Person',num2str(nn),'\a',num2str(i1),'t',num2str(i2),'.xlsx']);
% F = xlsread([frist_name(1,:),'\Person1','\SY',num2str(i),'.xlsx']);
cutoff = 3; % 截止频率
order = 10; % 滤波器阶数

% 低通滤波器设计
[b, a] = butter(order, cutoff / (fs / 2), 'low');

% 应用低通滤波到每个轴的数据
X = filtfilt(b, a, F);


V = X(:,1:3)*9.8/fs;
Q = X(:,4:6)*pi/180/fs;
V = [V(:,3),V(:,2),-V(:,1)];
Q = [Q(:,3),Q(:,2),-Q(:,1)];
avp0 = avpset([0;0;0], [0;0;0], [0;0;0]);  % 初始姿态（无意义），初始速度，初始位置
ins = insinit(avp0, 1/fs);

l = length(X);
N = floor(l/ft*2);
a32=0;
% Cbn = eye(3);

    nav = zeros(l, 9);Q3 = zeros(l,1);
    wvm =[Q(1,:) V(1,:)];
    [ins,phim,dvbm] = insupdate(ins, wvm);
    nav(ceil(1),:) = [q2att(ins.qnb)*180/pi; ins.vn; ins.pos];
    Q3(1) = nav(1,3);
    ins.vn(3) = 0;  ins.pos(3) = avp0(9); % 惯导天向通道发散，需要额外进行约束
for k = 2:l
    wvm =[Q(k,:) V(k,:)];
    [ins,phim,dvbm] = insupdate(ins, wvm);
    nav(ceil(k),:) = [q2att(ins.qnb)*180/pi; ins.vn; ins.pos];
    delta = nav(k,3) - nav(k-1,3);
    if delta >180
        Q3(k) = Q3(k-1) + delta - 360;
    else
        if delta <-180
```

```
            Q3(k) = Q3(k-1) + delta + 360;
        else
            Q3(k) = Q3(k-1) + delta;
        end
    end

    ins.vn(3) = 0;  ins.pos(3) = avp0(9); % 惯导天向通道发散，需要额外进行约束
end

A(31,i) = Q3(end);


for j = 1:N-1
    temp = X((j-1)*ft/2+1:(j+1)*ft/2,:);
    combined_acceleration = sqrt(temp(:, 1).^2+temp(:, 2).^2+temp(:, 3).^2);
    b = var(combined_acceleration);
    a32 = a32 + b;
end

temp = X(N*ft:end,:);
if ~isnan(temp)
    combined_acceleration = sqrt(temp(:, 1).^2+temp(:, 2).^2+temp(:, 3).^2);
    b = var(combined_acceleration);
    a32 = a32 + b;
    N = N+1;
end

    A(32,i) = a32/N;

    % 计算加速度计和陀螺仪各轴测量值的最大值和最小值

    A(1:6,i) = min(X);
    A(7:12,i) = max(X);
    % 计算加速度计和陀螺仪三轴数据之间的皮尔逊相关系数
    correlation_coefficients1 = corr([X(:,1),X(:,2),X(:,3)]);
    correlation_coefficients2 = corr([X(:,4),X(:,5),X(:,6)]);

    A(13,i) = correlation_coefficients1(2,1);
    A(14,i) = correlation_coefficients1(3,1);
    A(15,i) = correlation_coefficients1(3,2);
    A(16,i) = correlation_coefficients2(2,1);
    A(17,i) = correlation_coefficients2(3,1);
    A(18,i) = correlation_coefficients2(3,2);

    % 提取加速度计数据
    combined_acceleration = sqrt(X(:, 1).^2+X(:, 2).^2+X(:, 3).^2);

    % 计算最小值、最大值、方差、平均数、众数、四分位数、极差、四分位距、平均绝对离差以及信号能量
    A(19,i) = min(combined_acceleration);
    A(20,i) = max(combined_acceleration);
    A(21,i) = var(combined_acceleration);
    A(22,i) = mean(combined_acceleration);
    A(23,i) = median(combined_acceleration);
    A(24,i) = quantile(combined_acceleration, 0.25);
    A(25,i) = quantile(combined_acceleration, 0.75);
    A(26,i) = mode(combined_acceleration);
    A(27,i) = A(20,i) - A(19,i);
    A(28,i) = A(25,i) - A(24,i);
    A(29,i) = median(abs(combined_acceleration - A(22,i)));
    A(30,i) = sum(abs(fft(combined_acceleration)).^2) / length(combined_acceleration);
    end
    label = [label;i1*ones(5,1)];
end



B = zscore(A');

```

```
%     % 指定Excel文件的名称和要写入的工作表
%     fileName = ['Person',num2str(nn),'output.xlsx'];
%     sheetName = 'Sheet1';
%
%     % 使用xlswrite函数写入数据
%     xlswrite(fileName, B', sheetName);

    data = [data B'];
end
save('data.mat','data','-mat');
save('label.mat','label','-mat');
```

build_mode1.m 代码

build_model1.m

```
%% 问题1和问题2，用来完成模型训练，获取模型Mode1
clear
load('data.mat');load('label');
% 分割数据为训练集和测试集
% 这里我们假设前50个样本用于训练，后10个用于测试
X_train = data';
y_train = label;
X_test = data';
y_test = label;

n = size(data(1:32,:), 2);

% 随机划分，例如80%作为训练集，20%作为测试集
train_size = round(0.8 * n);
random_indices = randperm(n);
train_index = random_indices(1:train_size);
test_index = random_indices(train_size+1:end);

% 保存训练集和测试集的索引
save('train_index.mat', 'train_index');
save('test_index.mat', 'test_index');

% 使用训练集和测试集的索引来获取数据
X_train = data(:,train_index)';
y_train = label(train_index);
X_test = data(:,test_index)';
y_test = label(test_index);

% 创建SVM分类器
Model = fitcecoc(X_train, y_train, 'Learners', 'knn');
% % 创建决策树分类器
% Model = fitctree(X_train, y_train);
% %
% % % % 创建随机森林
% % % Model = TreeBagger(500, X_train, y_train);
% % %
% 使用测试集进行预测
y_pred = predict(Model, X_test);
% 计算准确率
accuracy = sum(y_pred == y_test) / length(y_test)

save('Model','Model');



% 使用pca函数进行主成分分析
[eigenvectors, scores] = pca(data', 'Algorithm','eig','Centered',false,'Rows','all','NumComponents',10);

% 提取第一主成分和第二主成分
reduced_features = scores * eigenvectors';

% 使用kmeans函数进行K聚类
```

```
k = 12; % 假设将数据分为3个群集
[cluster_labels, centroids] = kmeans(reduced_features, k);

% % 计算准确率
% accuracy = sum(cluster_labels == label) / length(label)

% 绘制图像
% 首先绘制每个群集的数据点
scatter(reduced_features(:, 1), reduced_features(:, 2));
hold on
% 然后绘制群集中心
scatter(centroids(:, 1), centroids(:, 2),'r','filled');

% 显示结果
title('K-means聚类结果');
xlabel('第一主成分');
ylabel('第二主成分');
```

text_mode1.m 代码

test_model.m

```
%% 问题1和问题2，用来完成数据预测和模型检验
clear
paths;
glvs;
A = zeros(32,60);B = zeros(32,60);data = [];
frist_name = ['test1';'test2';'test3';'test4';'test5'];
fs = 100; % 采样率为100 Hz
ft = 200; % 数据分割频率


for i1 = 1:3
    for i = 1:60
%% 数据初处理
F = xlsread([frist_name(1,:),'\Person',num2str(i1),'\SY',num2str(i),'.xlsx']);
cutoff = 3; % 截止频率
order = 10; % 滤波器阶数

% 低通滤波器设计
[b, a] = butter(order, cutoff / (fs / 2), 'low');

% 应用低通滤波到每个轴的数据
X = filtfilt(b, a, F);

    V = X(:,1:3)*9.8/fs;
    Q = X(:,4:6)*pi/180/fs;
    V = [V(:,3),V(:,2),-V(:,1)];
    Q = [Q(:,3),Q(:,2),-Q(:,1)];
    avp0 = avpset([0;0;0], [0;0;0], [0;0;0]);  % 初始姿态（无意义），初始速度，初始位置
    ins = insinit(avp0, 1/fs);

    l = length(X);
    N = floor(1/ft*2);
    a32=0;
    % Cbn = eye(3);

    nav = zeros(l, 9);Q3 = zeros(l,1);
    wvm =[Q(1,:) V(1,:)];
    [ins,phim,dvbm] = insupdate(ins, wvm);
    nav(ceil(1),:) = [q2att(ins.qnb)*180/pi; ins.vn; ins.pos];
    Q3(1) = nav(1,3);
    ins.vn(3) = 0;  ins.pos(3) = avp0(9); % 惯导天向通道发散，需要额外进行约束
for k = 2:l
    wvm =[Q(k,:) V(k,:)];
    [ins,phim,dvbm] = insupdate(ins, wvm);
    nav(ceil(k),:) = [q2att(ins.qnb)*180/pi; ins.vn; ins.pos];
    delta = nav(k,3) - nav(k-1,3);
    if delta >180
        Q3(k) = Q3(k-1) + delta - 360;
```

```
    else
        if delta <-180
            Q3(k) = Q3(k-1) + delta + 360;
        else
            Q3(k) = Q3(k-1) + delta;
        end
    end

    ins.vn(3) = 0;  ins.pos(3) = avp0(9); % 惯导天向通道发散，需要额外进行约束
end

A(31,i) = Q3(end);

for j = 1:N-1
    temp = X((j-1)*ft/2+1:(j+1)*ft/2,:);
    combined_acceleration = sqrt(temp(:, 1).^2+temp(:, 2).^2+temp(:, 3).^2);
    b = var(combined_acceleration);
    a32 = a32 + b;
end

temp = X(N*ft:end,:);
if ~isnan(temp)
    combined_acceleration = sqrt(temp(:, 1).^2+temp(:, 2).^2+temp(:, 3).^2);
    b = var(combined_acceleration);
    a32 = a32 + b;
    N = N+1;
end

    A(32,i) = a32/N;

    % 计算加速度计和陀螺仪各轴测量值的最大值和最小值
    A(1:6,i) = min(X);
    A(7:12,i) = max(X);

    % 计算加速度计和陀螺仪三轴数据之间的皮尔逊相关系数
    correlation_coefficients1 = corr([X(:,1),X(:,2),X(:,3)]);
    correlation_coefficients2 = corr([X(:,4),X(:,5),X(:,6)]);

    A(13,i) = correlation_coefficients1(2,1);
    A(14,i) = correlation_coefficients1(3,1);
    A(15,i) = correlation_coefficients1(3,2);
    A(16,i) = correlation_coefficients2(2,1);
    A(17,i) = correlation_coefficients2(3,1);
    A(18,i) = correlation_coefficients2(3,2);

    % 提取加速度计数据
    combined_acceleration = sqrt(X(:, 1).^2+X(:, 2).^2+X(:, 3).^2);

    % 计算最小值、最大值、方差、平均数、众数、四分位数、极差、四分位距、平均绝对离差以及信号能量
    A(19,i) = min(combined_acceleration);
    A(20,i) = max(combined_acceleration);
    A(21,i) = var(combined_acceleration);
    A(22,i) = mean(combined_acceleration);
    A(23,i) = median(combined_acceleration);
    A(24,i) = quantile(combined_acceleration, 0.25);
    A(25,i) = quantile(combined_acceleration, 0.75);
    A(26,i) = mode(combined_acceleration);
    A(27,i) = A(20,i) - A(19,i);
    A(28,i) = A(25,i) - A(24,i);
    A(29,i) = median(abs(combined_acceleration - A(22,i)));
    A(30,i) = sum(abs(fft(combined_acceleration)).^2) / length(combined_acceleration);

    end
    B = zscore(A');
    data = [data B'];
end

save('data.mat','data','-mat');
```

```
%     %归一化
% for i3 = 1:32
%     min_values = min(A(i3,:));
%     max_values = max(A(i3,:));
%     B(i3,:) = (A(i3,:) - min_values) / (max_values - min_values);
% end

load("Model.mat");
% 使用测试集进行预测
y_pred = predict(Model, data');

% % 计算准确率
% accuracy = sum(y_pred == y_test) / length(y_test);




% % 使用pca函数进行主成分分析
% [eigenvectors, scores] = pca(data, 'Algorithm','eig','Centered',false,'Rows','all','NumComponents',2);
%
% % 提取第一主成分和第二主成分
% reduced_features = scores * eigenvectors';
%
% % 使用kmeans函数进行K聚类
% k = 12; % 假设将数据分为3个群集
% [cluster_labels, centroids] = kmeans(reduced_features, k);
%
% % 绘制图像
% % 首先绘制每个群集的数据点
% scatter(reduced_features(:, 1), reduced_features(:, 2));
% hold on
% % % 然后绘制群集中心
% % scatter(centroids(:, 1), centroids(:, 2));
%
% % 显示结果
% title('K聚类结果');
% xlabel('第一主成分');
% ylabel('第二主成分');
```

附录 B：问题 3 代码

首先利用 get_date3 程序完成问题 3 获取数据部分，然后建立模型 build_mode3，最终检验模型 text_mode3。

get_date3.m

get_data3.m

```
%%问题3 用来读取数据，完成数据预处理和获取数据特征矩阵data
clear
paths;
glvs;
A = zeros(32,50);B = zeros(32,50);
frist_name = ['test1';'test2';'test3';'test4';'test5'];
fs = 100; % 采样率为100 Hz
ft = 200; % 数据分割频率

% figure
test4 = xlsread([frist_name(4,:),'.xlsx']);
person = test4(4:13,2:4);
for i = 1:10
    person_data(5*(i-1)+1:5*i,:) = person(i,:).*ones(5,1);
end

label = [];data = [];
for nn = 1:12
```

```
for i1 = 4:13
    for i2 = 1:5
    i=(i1-4)*5+i2;
%% 数据初处理
F = xlsread([frist_name(2,:),'\Person',num2str(i1),'\a',num2str(nn),'t',num2str(i2),'.xlsx']);

cutoff = 3; % 截止频率
order = 10; % 滤波器阶数

% 低通滤波器设计
[b, a] = butter(order, cutoff / (fs / 2), 'low');

% 应用低通滤波到每个轴的数据
X = filtfilt(b, a, F);




V = X(:,1:3)*9.8/fs;
Q = X(:,4:6)*pi/180/fs;
V = [V(:,3),V(:,2),-V(:,1)];
Q = [Q(:,3),Q(:,2),-Q(:,1)];
avp0 = avpset([0;0;0], [0;0;0], [0;0;0]);  % 初始姿态（无意义），初始速度，初始位置
ins = insinit(avp0, 1/fs);

l = length(X);
N = floor(l/ft*2);
a32=0;

    nav = zeros(1, 9);Q3 = zeros(1,1);
    wvm =[Q(1,:) V(1,:)];
    [ins,phim,dvbm] = insupdate(ins, wvm);
    nav(ceil(1),:) = [q2att(ins.qnb)*180/pi; ins.vn; ins.pos];
    Q3(1) = nav(1,3);
    ins.vn(3) = 0;  ins.pos(3) = avp0(9); % 惯导天向通道发散，需要额外进行约束
for k = 2:l
    wvm =[Q(k,:) V(k,:)];
    [ins,phim,dvbm] = insupdate(ins, wvm);
    nav(ceil(k),:) = [q2att(ins.qnb)*180/pi; ins.vn; ins.pos];
    delta = nav(k,3) - nav(k-1,3);
    if delta >180
        Q3(k) = Q3(k-1) + delta - 360;
    else
        if delta <-180
            Q3(k) = Q3(k-1) + delta + 360;
        else
            Q3(k) = Q3(k-1) + delta;
        end
    end

    ins.vn(3) = 0;  ins.pos(3) = avp0(9); % 惯导天向通道发散，需要额外进行约束
end

A(31,i) = Q3(end);

for j = 1:N-1
    temp = X((j-1)*ft/2+1:(j+1)*ft/2,:);
    combined_acceleration = sqrt(temp(:, 1).^2+temp(:, 2).^2+temp(:, 3).^2);
    b = var(combined_acceleration);
    a32 = a32 + b;
end

temp = X(N*ft:end,:);
if ~isnan(temp)
    combined_acceleration = sqrt(temp(:, 1).^2+temp(:, 2).^2+temp(:, 3).^2);
    b = var(combined_acceleration);
    a32 = a32 + b;
```

```
    N = N+1;
end

    A(32,i) = a32/N;


    % 计算加速度计和陀螺仪各轴测量值的最大值和最小值

    A(1:6,i) = min(X);
    A(7:12,i) = max(X);
    % 计算加速度计和陀螺仪三轴数据之间的皮尔逊相关系数
    correlation_coefficients1 = corr([X(:,1),X(:,2),X(:,3)]);
    correlation_coefficients2 = corr([X(:,4),X(:,5),X(:,6)]);

    A(13,i) = correlation_coefficients1(2,1);
    A(14,i) = correlation_coefficients1(3,1);
    A(15,i) = correlation_coefficients1(3,2);
    A(16,i) = correlation_coefficients2(2,1);
    A(17,i) = correlation_coefficients2(3,1);
    A(18,i) = correlation_coefficients2(3,2);

    % 提取加速度计数据
    combined_acceleration = sqrt(X(:, 1).^2+X(:, 2).^2+X(:, 3).^2);

    % 计算最小值、最大值、方差、平均数、众数、四分位数、极差、四分位距、平均绝对离差以及信号能量
    A(19,i) = min(combined_acceleration);
    A(20,i) = max(combined_acceleration);
    A(21,i) = var(combined_acceleration);
    A(22,i) = mean(combined_acceleration);
    A(23,i) = median(combined_acceleration);
    A(24,i) = quantile(combined_acceleration, 0.25);
    A(25,i) = quantile(combined_acceleration, 0.75);
    A(26,i) = mode(combined_acceleration);
    A(27,i) = A(20,i) - A(19,i);
    A(28,i) = A(25,i) - A(24,i);
    A(29,i) = median(abs(combined_acceleration - A(22,i)));
    A(30,i) = sum(abs(fft(combined_acceleration)).^2) / length(combined_acceleration);
    end
    label = [label;il*ones(5,1)];
end

%     %归一化
% for i3 = 1:32
%     min_values = min(A(i3,:));
%     max_values = max(A(i3,:));
%     B(i3,:) = (A(i3,:) - min_values) / (max_values - min_values);
% end

%     % 指定Excel文件的名称和要写入的工作表
%     fileName = ['Person',num2str(nn),'output.xlsx'];
%     sheetName = 'Sheet1';
%
%     % 使用xlswrite函数写入数据
%     xlswrite(fileName, B', sheetName);

    C = [A' person_data];
    B = zscore(C);
    data = [data B'];
end
save('data.mat','data','-mat');
save('label.mat','label','-mat');
```

build_mode3.m

build_model3.m

```
%% 问题3，用来完成模型训练，获取模型Mode3
clear
load('data.mat');load('label');
% 分割数据为训练集和测试集
% % 这里我们假设前50个样本用于训练，后10个用于测试
% X_train = data';
% y_train = label;
% X_test = data';
% y_test = label;

y_pred = zeros(12,10);

for i = 1:12
    temp1 = [data(1:32,(i-1)*50+1:i*50);data(33:35,(i-1)*50+1:i*50)];
    temp2 = label((i-1)*50+1:i*50);
%     n = size(temp1, 2);
% % 随机划分，例如80%作为训练集，20%作为测试集
% train_size = round(0.8 * n);
% random_indices = randperm(n);
% train_index = random_indices(1:train_size);
% test_index = random_indices(train_size+1:end);

% % 保存训练集和测试集的索引
% save('train_index.mat', 'train_index');
% save('test_index.mat', 'test_index');
k1=0;k2=0;
for j = 1:50
    if mod(j,5) == 0
        k1= k1+1;
        test_index(k1) = j;
    else
        k2 = k2+1;
        train_index(k2) = j;
    end
end


% 使用训练集和测试集的索引来获取数据
X_train = temp1(:,train_index)';
y_train = temp2(train_index);
X_test = temp1(:,test_index)';
y_test = temp2(test_index);

% 创建SVM分类器
Mode3 = fitcecoc(X_train, y_train, 'Learners', 'linear');
% % 创建决策树分类器
% Model = fitctree(X_train, y_train);
% %
% % % % 创建随机森林
% % % Model = TreeBagger(500, X_train, y_train);
% % %
% 使用测试集进行预测
y_pred(i,:) = predict(Mode3, X_test);
% 计算准确率
accuracy(i) = sum(y_pred(i,:)' == y_test) / length(y_test);

Mode3name = ['Mode3',num2str(i)];
save(Mode3name,'Mode3');

end

```

```
result_accuracy =  mean(accuracy)

% for i = 1:10
%     k = mode(y_pred(:,(i-1)*5+1:i*5),'all');
%     if  k == i+3
%         result(i) = 1;
%     else
%         result(i) = 0;
%     end
% end
% accuracy = sum(result)/10

% % 使用pca函数进行主成分分析
% [eigenvectors, scores] = pca(data', 'Algorithm','eig','Centered',false,'Rows','all','NumComponents',10);
%
% % 提取第一主成分和第二主成分
% reduced_features = scores * eigenvectors';
%
% % 使用kmeans函数进行K聚类
% k = 12; % 假设将数据分为3个群集
% [cluster_labels, centroids] = kmeans(reduced_features, k);
%
% % % 计算准确率
% % accuracy = sum(cluster_labels == label) / length(label)
%
% % 绘制图像
% % 首先绘制每个群集的数据点
% scatter(reduced_features(:, 1), reduced_features(:, 2));
% hold on
% % 然后绘制群集中心
% scatter(centroids(:, 1), centroids(:, 2),'r','filled');
%
% % 显示结果
% title('K-means聚类结果');
% xlabel('第一主成分');
% ylabel('第二主成分');
```

text_mode3.m

test_model3.m*

```
%% 问题3，用来完成数据预测和模型检验
clear
paths;
glvs;
A = zeros(32,5);B = zeros(32,5);data = [];
frist_name = ['test1';'test2';'test3';'test4';'test5'];
fs = 100; % 采样率为100 Hz
ft = 200; % 数据分割频率


for i1 = 1:12
    for i = 1:5
%% 数据初处理
F = xlsread([frist_name(5,:),'\unknow',num2str(i),'\a',num2str(i1),'t1.xlsx']);
cutoff = 3; % 截止频率
order = 10; % 滤波器阶数

% 低通滤波器设计
[b, a] = butter(order, cutoff / (fs / 2), 'low');

% 应用低通滤波到每个轴的数据
X = filtfilt(b, a, F);

    V = X(:,1:3)*9.8/fs;
    Q = X(:,4:6)*pi/180/fs;
    V = [V(:,3),V(:,2),-V(:,1)];
    Q = [Q(:,3),Q(:,2),-Q(:,1)];
    avp0 = avpset([0;0;0], [0;0;0], [0;0;0]);  % 初始姿态（无意义），初始速度，初始位置
```

```
    ins = insinit(avp0, 1/fs);

    l = length(X);
    N = floor(1/ft*2);
    a32=0;
    % Cbn = eye(3);

    nav = zeros(1, 9);Q3 = zeros(1,1);
    wvm =[Q(1,:) V(1,:)];
    [ins,phim,dvbm] = insupdate(ins, wvm);
    nav(ceil(1),:) = [q2att(ins.qnb)*180/pi; ins.vn; ins.pos];
    Q3(1) = nav(1,3);
    ins.vn(3) = 0;  ins.pos(3) = avp0(9); % 惯导天向通道发散，需要额外进行约束
for k = 2:l
    wvm =[Q(k,:) V(k,:)];
    [ins,phim,dvbm] = insupdate(ins, wvm);
    nav(ceil(k),:) = [q2att(ins.qnb)*180/pi; ins.vn; ins.pos];
    delta = nav(k,3) - nav(k-1,3);
    if delta >180
        Q3(k) = Q3(k-1) + delta - 360;
    else
        if delta <-180
            Q3(k) = Q3(k-1) + delta + 360;
        else
            Q3(k) = Q3(k-1) + delta;
        end
    end

    ins.vn(3) = 0;  ins.pos(3) = avp0(9); % 惯导天向通道发散，需要额外进行约束
end

A(31,i) = Q3(end);

for j = 1:N-1
    temp = X((j-1)*ft/2+1:(j+1)*ft/2,:);
    combined_acceleration = sqrt(temp(:, 1).^2+temp(:, 2).^2+temp(:, 3).^2);
    b = var(combined_acceleration);
    a32 = a32 + b;
end

temp = X(N*ft:end,:);
if ~isnan(temp)
    combined_acceleration = sqrt(temp(:, 1).^2+temp(:, 2).^2+temp(:, 3).^2);
    b = var(combined_acceleration);
    a32 = a32 + b;
    N = N+1;
end

    A(32,i) = a32/N;

    % 计算加速度计和陀螺仪各轴测量值的最大值和最小值
    A(1:6,i) = min(X);
    A(7:12,i) = max(X);

    % 计算加速度计和陀螺仪三轴数据之间的皮尔逊相关系数
    correlation_coefficients1 = corr([X(:,1),X(:,2),X(:,3)]);
    correlation_coefficients2 = corr([X(:,4),X(:,5),X(:,6)]);

    A(13,i) = correlation_coefficients1(2,1);
    A(14,i) = correlation_coefficients1(3,1);
    A(15,i) = correlation_coefficients1(3,2);
    A(16,i) = correlation_coefficients2(2,1);
    A(17,i) = correlation_coefficients2(3,1);
    A(18,i) = correlation_coefficients2(3,2);

    % 提取加速度计数据
    combined_acceleration = sqrt(X(:, 1).^2+X(:, 2).^2+X(:, 3).^2);
```

```
    % 计算最小值、最大值、方差、平均数、众数、四分位数、极差、四分位距、平均绝对离差以及信号能量
    A(19,i) = min(combined_acceleration);
    A(20,i) = max(combined_acceleration);
    A(21,i) = var(combined_acceleration);
    A(22,i) = mean(combined_acceleration);
    A(23,i) = median(combined_acceleration);
    A(24,i) = quantile(combined_acceleration, 0.25);
    A(25,i) = quantile(combined_acceleration, 0.75);
    A(26,i) = mode(combined_acceleration);
    A(27,i) = A(20,i) - A(19,i);
    A(28,i) = A(25,i) - A(24,i);
    A(29,i) = median(abs(combined_acceleration - A(22,i)));
    A(30,i) = sum(abs(fft(combined_acceleration)).^2) / length(combined_acceleration);


    end
    B = zscore(A');
    data = [data B'];
end

%     %归一化
% for i3 = 1:32
%     min_values = min(A(i3,:));
%     max_values = max(A(i3,:));
%     B(i3,:) = (A(i3,:) - min_values) / (max_values - min_values);
% end

load("Model.mat");
y_pred = zeros(5,12);
for i = 1:12
    Modelname = ['Model',num2str(i)];
    load(Modelname);
    y_pred(:,i) = predict(Model, data(:,(i-1)*5+1:i*5)');
end


% 使用测试集进行预测


% % 计算准确率
% accuracy = sum(y_pred == y_test) / length(y_test);


% % 使用pca函数进行主成分分析
% [eigenvectors, scores] = pca(data, 'Algorithm','eig','Centered',false,'Rows','all','NumComponents',2);
%
% % 提取第一主成分和第二主成分
% reduced_features = scores * eigenvectors';
%
% % 使用kmeans函数进行K聚类
% k = 12; % 假设将数据分为3个群集
% [cluster_labels, centroids] = kmeans(reduced_features, k);
%
% % 绘制图像
% % 首先绘制每个群集的数据点
% scatter(reduced_features(:, 1), reduced_features(:, 2));
% hold on
% % % 然后绘制群集中心
% % scatter(centroids(:, 1), centroids(:, 2));
%
% % 显示结果
% title('K聚类结果');
% xlabel('第一主成分');
% ylabel('第二主成分');
```