1. **A Multi-Featured Approach for Wearable Sensor-based Human Activity Recognition**
2. **A Hierarchical Meta-Classifier for Human Activity Recognition**

2016 15th IEEE International Conference on Machine Learning and Applications

**数据采集、预处理及实验**：

活动类型：5组，分别为静止，步行，跑步，上楼梯，下楼梯；再细分为23个子类。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 活动组别 | 编号 | 活动类型 | 时间/距离 | 采集人数 |
| 步行 | 1 | 平地跑步机1mph | 3min | 29 |
| 2 | 平地跑步机2mph | 3min | 21 |
| 3 | 平地跑步机3mph | 3min | 28 |
| 4 | 坡度5%跑步机3mph | 3min | 29 |
| 跑步 | 5 | 平地跑步机5mph | 3min | 21 |
| 6 | 平地跑步机6mph | 3min | 34 |
| 7 | 坡度5%跑步机6mph | 3min | 26 |
| 静止 | 8 | 坐，整理/堆叠衣服 | 3min | 74 |
| 9 | 站立/谈话时手动（坐立不安） | 3min | 77 |
| 10 | 1min刷牙+1min梳头 | 2min | 77 |
| 11 | 开车 |  | 21 |
| 步行 | 12 | 硬地步行-运动鞋 | 400m | 76 |
| 13 | 硬地步行-运动鞋 手在口袋中 | 100m | 33 |
| 14 | 硬地步行-运动鞋 负重8lb物品 | 100m | 30 |
| 15 | 硬地步行-运动鞋 手持电话 | 100m | 24 |
| 16 | 硬地步行-运动鞋 手持满杯咖啡 | 100m | 26 |
| 17 | 地毯步行-高跟鞋/舞鞋 | 100m | 70 |
| 18 | 草坪赤脚步行 | 134m | 20 |
| 19 | 不平的土地-运动鞋 | 107m | 23 |
| 20 | 上坡度5%山坡，高跟鞋/舞鞋 | 58.5m\*2次 | 27 |
| 21 | 下坡度5%山坡，高跟鞋/舞鞋 | 58.5m\*2次 | 26 |
| 上楼梯 | 22 | 上楼梯，5层 | 5层\*2次 | 77 |
| 下楼梯 | 23 | 下楼梯，5层 | 5层\*2次 | 77 |

招募77名志愿者，18-64岁，55男，22女，腰部三轴加速计（阈值-6g-6g），采样频率100Hz；

每人均完成如下活动：站立，谈话时手动（坐立不安），1min刷牙，1min梳头，某些形式的硬地或地毯行走，上下楼梯。此外三种跑步机运动，三种其他运动随机分配。

**特征提取**：246特征，2s时间窗口

时域特征：针对三轴以及对应的一阶差分、合向量，分别选取均值，最大值，最小值，标准差，穿过中点值（信号低于中值到超过中值，或相反，总次数），10/25/50/75/90分位数，各轴间及一阶差分间的相关系数；

频域特征：各轴的幅度magnitude，一阶差分及合向量的幅度；

小波分析特征：针对xyz各轴，1-5级小波分解系数，同样包括均值，最大值，最小值，标准差，穿过中点值，10/25/50/75/90分位数；

样本统计特征：年龄，体重，身高，性别，BMI指数

**特征选择：**

相关系数法，relief-based，专家判断(基于多种特征选择方法，主观判断选择42个特征)；

Relief-based: 特征对于从属于不同类别的相似点的区分能力。随机选取个例，搜索和该个例最近但从属于不同类别的个例。依据对相似点（miss,hit）的区分能力，对特征赋予权值。通过对每条数据进行迭代更新。选取能够携带更多信息的特征；缺点是没有考虑特征的独立性，不能有效剔除冗余特征。

相关系数法：特征集合排序而非单独特征。依据特征的相关性，对特征集合打分。综合考虑了区分不同类别的能力，以及不同特征的内在关联。通常可选取到和类别高度相关，而组内各特征相关性弱的特征集合。

四个特征组别(全体特征)分别基于决策树，NB，KNN（K=5），随机森林4种分类器分类，各特征集合表现相似，随机森林分类器效果最好。

**分类过程：**

分层分类器：高层分类器，区分5个组别；底层分类器区分更细致的组内活动类别；

弊端在于高层的错分个例会传导到底层分类器，因而高层分类器精度必须足够高。

87943条记录，9:1 训练、测试集，因数据集为非平衡数据，按类型比例划分数据。

5个随机森林分类器，3个层级

第一层，活动分为相对静止，非静止两类， 97.899%;

第二层，活动分组，静止为第一层，非静止细分为步行，跑步及上下楼梯，94%;

第三层，细分为23个子类别，整体准确率87%，静止，步行和跑步分别为94%，89%，97%，上下楼梯因只有一个子类别，准确率为81%，75.5%（归因为个体差异，如年龄/体重）。

**研究亮点**：1）数据集大，活动类型更多；2）采用hierarchical分层策略，相比于单层分类器分类效果提升；3）细分类别更多，相比于类似研究，更具有实际意义。

1. **Activity Recognition on Smartphones: Efficient Sampling Rates and Window Sizes**

**时间窗口尺寸+采样比率——>CPU占用率+准确率**

**数据采集**：采样频率80Hz，5s时间窗；

7种类型活动：步行（107），慢跑（102），坐（104），躺（110），站立（106），上楼梯（24），下楼梯（19）

3名志愿者，不同身体情况（超重，高个，矮个），手机置于裤子前口袋，每种活动2分钟。

特征提取：61维特征

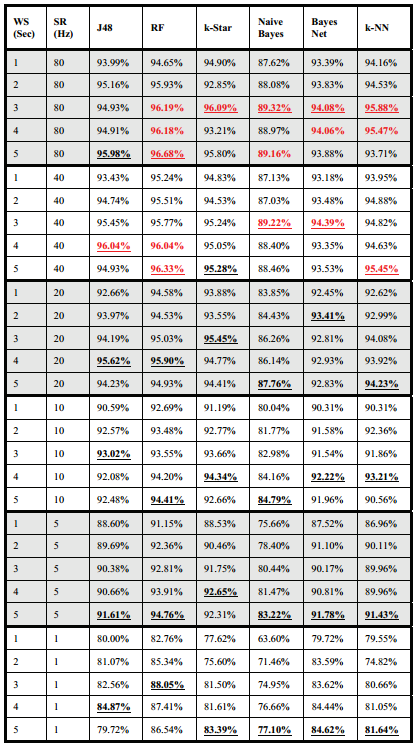
数据降采样：40Hz，20Hz，10Hz，5Hz，1Hz；

时间窗口：4s，3s，2s，1s

六种分类器：C4.5，k-Star，朴素贝叶斯，贝叶斯网络，随机森林，KNN；10折交叉验证

CPU占用率：iPhone4 CPU 时钟频率 800Mhz，在iPod Touch 6 验证

**实验结果：<时间窗，采样频率>**

**1 采样频率+时间窗——>准确率**

采样频率高，准确率高；对于J48分类器，<4,40>准确率反而高于<4,80>;

时间窗3-5s，通常有最佳结果；采样频率降低后，更大的时间窗准确率越高；

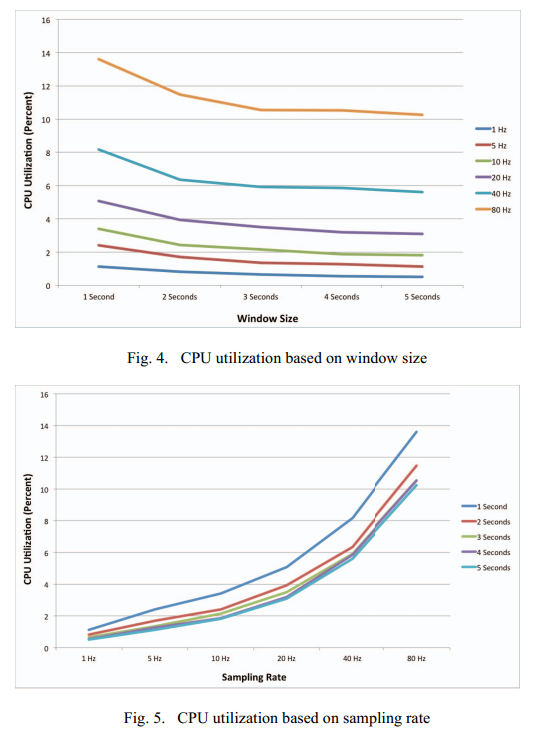
**2 采样频率+时间窗——>CPU占用率**

当时间窗尺寸增加，CPU占用率下降；当时间窗超过3s,CPU占用率减速很缓，可以忽视；而每种分类器的最佳结果基本都在3-5s，因而这个结果很重要。

采样频率对CPU占用率影响更大。

**3 采样频率+时间窗 有效**

尽可能高的准确率+低的CPU占用

有效的时间窗尺寸是3-5s; 有效的采样频率是5-20Hz。

1. **Simple and Complex Activity Recognition Through Smart Phones-2012**

简单活动：自行车，上楼梯，开车，躺，跑步，坐，站立，步行，手机离身

复杂活动新增：清洗(厨房柜台和水槽)，做饭（模拟微波炉烧热水，倒水），取药（从橱柜取出一周药品），扫地，洗手（香皂，厨房水槽），浇水（两个房间的三种植物）

特征提取：时间窗1,2,4,8,12,16s；半重叠，每个时间点记录贡献给两个个例。

分类器：MLP, NB , 贝叶斯网络，决策树，Best-First Tree，K-star

结果：简单活动准确率高：MLP，2s时间窗；时间窗对简单活动分类结果影响很小；

复杂活动准确率低，略高于50%，两种活动混合时，简单活动依然保持高准确率。

复杂活动分类不能单独应用于实践，不过可以考虑和其他系统结合。

1. **A Study on Human Activity Recognition Using Accelerometer Data from Smartphones-2014**

数据采集：2男，2女，29-33，手/兜

活动：跑步，慢走，快走，跳舞，上楼，下楼

采样频率：100Hz，onesample10ms，共收集79573条数据

特征提取：128samples，1.28s，半重叠；

抽取24维特征：均值，最小，最大，标准差，RMS，APF(峰值个数)，各轴相关系数等

分类器：MLP，随机森林，LMT，SVM，LOGIT，LOGITBoost

MLP准确率最高：手中89.48%，兜里89.72%；随机森林，SVM其次

组合学习：分类器组合，多数投票

手里：MLP+LogitBoost+SVM: 91.15%；兜里：MLP+随机森林+LOGIT，90.34%

特征选择相关资料总结：

1. [米筐科技（RiceQuant）策略研究报告：特征选择方法探析—沪深300指数的集成特征选择和聚类分析](https://zhuanlan.zhihu.com/p/21406355)
2. [机器学习中，有哪些特征选择的工程方法？](https://www.zhihu.com/question/28641663)
3. [干货：结合Scikit-learn介绍几种常用的特征选择方法](http://dataunion.org/14072.html)
4. [特征选择方法总结](http://blog.csdn.net/lihaitao000/article/details/51213563?hmsr=toutiao.io&utm_medium=toutiao.io&utm_source=toutiao.io)