

机器学习在数据分析预测中的应用之人类行为分类

摘要：人类活动行为的识别和分类是机器学习领域的一个热点话题。通过对来自于随身佩戴的传感器的数据进行处理和分析，用机器学习的方式建立一个可以对不同行为进行区分的模型，医生和软件开发人员可以了解个人的生活方式和活动选择，并据此给出合理的生活建议。本小组使用PAMAP2数据集，用多种机器学习方法对该分类问题进行建模处理，并逐步提高模型的准确度和稳定性，最终得到了具有良好效能的模型，并对不同机器学习算法的结果进行了对比。

关键词：机器学习 人类活动行为 数据处理 数据分析预测

1 背景简介

活动行为识别是多种医疗保健应用程序中的一项重要任务。通过持续地检测和分析用户的活动行为，应用程序可以自动为患者和医生提供合理的个性化推荐。还有一些应用程序用于消费产品，比如智能手机和手表中记录健康数据的应用等。常见的消费产品中通常附带惯性测量单元(Inertial Measurement Unit, IMU)。该测量单元内含温度计、加速度计和陀螺仪传感器，通过其提供的信息，运用机器学习技术可以训练一个活动分类器，使医生和应用程序开发人员了解个人的生活方式和活动选择，并据此给出合理的生活建议。在本项目中，我们运用了多种机器学习的算法，希望训练出一个训练成本低、准确性高、稳定性好、可信度强的机器学习模型。我们综合应用了数据分析预测的各类训练和评价方法，探索了多种数据预处理方式，研究了逻辑回归、K邻近、各类决策树技术、支持向量机和神经网络技术的应用，对多种模型进行了多角度对比，并在多种可选参数组合中选取了合适的参数训练模型。最终我们获得了一个相对稳定和健壮的分类模型。

2 相关工作

基于有监督学习的基本理论，利用传感器数据进行人类活动行为分类是一个热门的研究主题。通过调研，我们发现类似工作的文章主要运用的机器学习方法有三类：朴素贝叶斯分类、支持向量机(SVM)/决策树分类和神经网络分类。使用朴素贝叶斯方法作为人类行为活动的分类器是有趣且聪明的，因为它通常用于文本分类。其中一篇使用朴素贝叶斯分类器的文章是Long, Yin和Aarts在2009年的论文[1]。该文章使用来自多个志愿者的传感器数据进行多类活动分类，并使用交叉验证的方式评估模型（并将朴素贝叶斯模型同包括决策树在内的其他模型进行比较）。然而，他们的模型准确率仅有80%左右。[Parkka, 2006][2]是一篇研究决策树的论文。该研究使用了交叉验证的方式评估了普通决策树在人类行

为分类中的效果，并使用了基尼指数作为损失。研究人员利用他们的专家知识和对传感器数据的预分析创建了一个“自定义决策树”。不过，文中使用的数据集极不平衡（一个类别占有所有数据的比例甚至达到了60%），这可能会影响测试性能。Youngwook Kim和Hao Ling的论文[3]中讨论了一种通过多普勒雷达(Doppler Radar)获取的数据来识别人类活动行为的方法，他们基于决策树和SVM模型选择出了一个最优化的模型。该论文中的最终分类正确率达到了90%，然而，文中数据的特征与我们希望处理的来自IMU的数据特征有着本质上的不同。近年来最先进的人类行为分类器常常运用神经网络模型。[Hammerla, 2016][4]将卷积神经网络(CNN)与滑动窗口结合起来，以查看用户一段时间内的活动，并最终获得了95%的分类准确率。

以上相关工作使用了多种模型对人类行为活动进行分类。他们采用的数据集通常是来源于传感器的数据，但是数据集容量、数据集内容和数据集中的特征数量都参差不齐，并且模型数量有限，没有多角度对比多种模型。

在本文的第3部分中将详细介绍用于本研究项目的基于IMU传感器采样的数据集PAMAP2，并窥探该数据集的基本特征；在本文的第4部分中将介绍本文使用的多种分类模型的基本原理；在本文的第5部分中将介绍模型训练和评估的基本方法；在第6部分中将介绍本项目的实验步骤、模型训练过程和结果分析；在第7部分中将对本项目进行总结，并归纳启示。

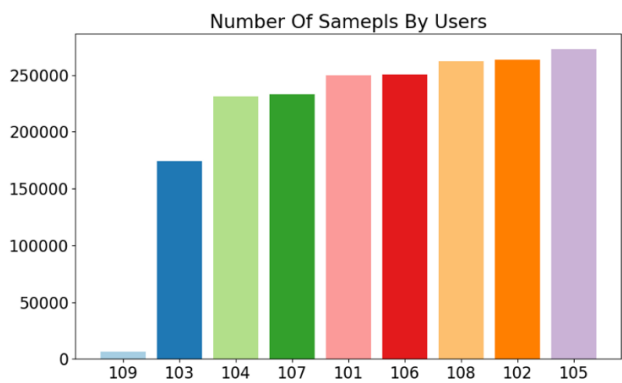
3 数据集介绍和数据探索

本项目使用来自UCI机器学习数据集存储库的PAMAP2数据集。该数据集中包含了九名志愿者的心率，以及他们分别佩戴于胸部、手腕与脚踝出的ICU单元所测量的数据，每个IMU以100hz的频率测量了体温、3轴加速度、3轴角速度和3轴磁力计的数据。数据集中总共有190万个数据点，每个数据点包含52个特征。我们的目标特征为志愿者当前的活动类型，包含行走、躺、上楼、下楼等12种不同的活动类型，每个类型对应一个

activity_ID。在实验中我们通过数据集中的体温、心率、加速度等描述性特征，以志愿者的活动类型为目标特征来构建模型，进行拟合与预测。

3.1 数据集概览

在PAMAP2数据集中，每位志愿者的测量数据数量并不平均，编号为109的志愿者数据数量远小于其他志愿者。但根据实验目的，我们希望的是对于任何人做出的动作，都能够相应地正确判断其当下的活动类型，并不考虑个体之间的差异。总数据量十分庞大，故在后续实验中应当将所有数据整合到一起进行分析预测。整合后的数据集共有超过190万条样例，样例数目十分庞大，如果不加以额外处理就直接进行模型训练是十分困难的。



3.2 缺失值

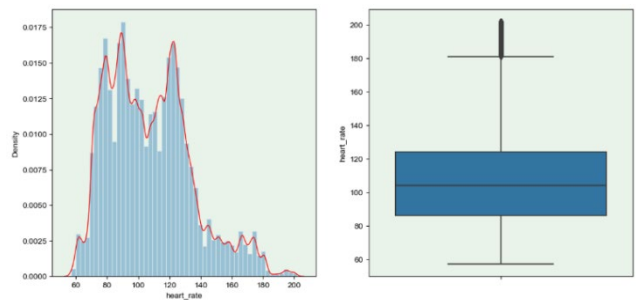
在数据集中，IMU以100Hz的频率测量相应的数据，而心率以9Hz的频率被检测记录，这导致了数据集中接近九成的数据点的心率值都是NaN。根据人体生理特征随时间的变化规律，我们使用线性填充的方法填充所有的缺失值。

time_stamp	activity_id	heart_rate	hand_temperature	hand_3D_acceleration_16_x
5.64	0	NaN	33.000	2.791430
5.65	0	NaN	33.000	2.860860
5.66	0	100.0	33.000	2.809970
5.67	0	NaN	33.000	2.771380
5.68	0	NaN	33.000	2.767100
...
4475.59	0	NaN	29.125	-0.040828
4475.60	0	NaN	29.125	-0.004109
4475.61	0	NaN	29.125	-0.073003
4475.62	0	NaN	29.125	0.189650
4475.63	0	134.0	29.125	-0.033339

3.3 异常值

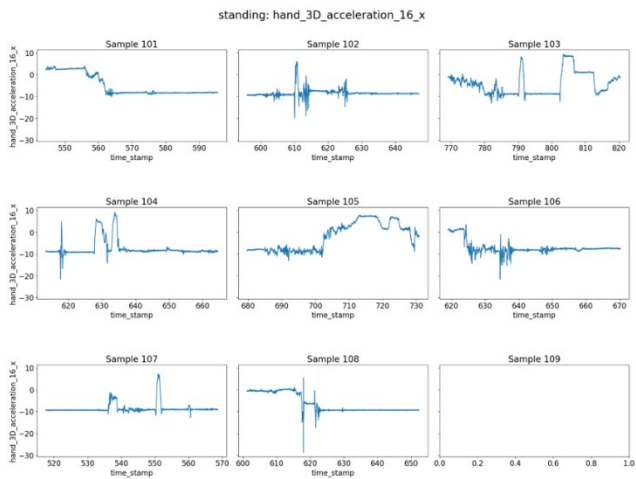
对每个特征的数据绘制了相应的箱型图，在箱型图中可以查看每个描述性特征是否存在偏离的异常值。每

个特征在箱型图中都有偏离的数据，但是在较大较小值处，特征数据分布十分密集。结合实际情况，我们认为这样的数据是合理的，不予删去。



3.4 数据冗余

通过对传感器数据以时间为轴分别绘制了对应九名志愿者的折线图。其中可以发现，每位志愿者的折线图在实验早期与实验末期的图像与其他时段相比十分平滑，几乎与时间横轴平行。我们推测这是志愿者在实验开始与结束的期间内做出的无效的动作，因此这部分数据很可能含有冗余信息，不利于准确分类。



3.5 归一化

数据集中，不同特征的范围十分广泛，平均值和标准差等数据特征差异也较大。这都导致在后续的处理中可能会出现例如梯度消失或梯度爆炸等问题。因此，采用归一化的方式进行数据预处理是十分必要的。常见的数据归一化方法主要有3种。其一是数据标准化(Standard Scaler)，将原数据转化为均值为0、方差为1的新数据；其二是数据范围化(MinMax Scaler)，将原数据压缩到一个给定范围之内(常见的范围如[-1, 1])；其三是Yeo-Johnson Transformer，通过一系列参数化的单调变换将数据尽可能地映射到高斯分布，有利于模型建构。在后续实验中，我们分别尝试了不同的归一化方式，选取其中效果最好的一种。

4 模型简介

4.1 含正则项的逻辑回归

加入正则化可以减小逻辑回归过拟合的概率。其损失函数如下所示：

$$J(\theta) = \sum_{i=0}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \frac{1}{C} (\|\theta\|_2)^2$$

为了使模型对过拟合具有更强的鲁棒性，将L2正则化与C参数正则化强度成反比。为了选择C的最优值，模型在C值的范围内进行独立训练。在验证集上产生最高准确性的C值被选择并在测试集上进行测试。

4.2 K-NN

K邻近是一种轻量级的学习算法。通过使用如下所示的不同的计算距离方式——欧氏距离与曼哈顿距离，来进行不同地拟合：

$$Euclidean(a, b) = \sqrt{\sum_{i=1}^m (a[i] - b[i])^2}$$

$$Manhattan(a, b) = \sum_{i=1}^m abs(a[i] - b[i])$$

为了选择K的最优值，模型也对不同的K值进行了训练，找到最高准确性的K值。该模型最终的计算公式为：

$$M_k(q) = \arg \max \sum_{i=1}^k \frac{1}{dist(q, d^{(i)})^2} \times \delta(t_i, l)$$

其中 $l \in level(t)$ 。

4.3 支持向量机(SVM)

含有软间隔的支持向量机模型可以构建一个非线性的模型。SVM的模型公式如下：

$$\min_{y, w, b} = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \zeta_i$$

$$\text{s.t. } y^{(i)}(w^T x^{(i)} + b) \geq 1 - \zeta_i, i = 1, 2, \dots, m$$

$$\zeta_i \geq 0, i = 1, 2, \dots, m$$

其决策函数的定义为：

$$\text{sgn} \left(\sum_{i=1}^n y_i \alpha_i K(x_i, x) + \rho \right)$$

其中sgn函数的定义为：

$$\text{sgn}(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}$$

为了得到最优的C值，我们在模型中对不同的C值进行了训练。同时，我们也尝试了使用不同的核方法，如线性核、高斯核等。

4.4 AdaBoost

AdaBoost模型实际上就是对多个基分类器进行训练。每一轮都会对其中的一个基分类器进行训练，其中根据前一轮的训练结果赋予样本点不同的权值。训练完成后根据每个基分类器的误差率给予分类器不同的权值并将其线性组合在一起形成最终的分类器模型。在构建AdaBoost模型中，我们选择了决策树作为基本的分类器。为了选择出实验效果最好的参数值，我们在实验过程中调整了基分类器的个数，并且在决策树的基分类器中分别使用基尼系数与信息熵作为不纯度的指标以帮助寻找最佳节点与最佳分支。

4.5 随机森林

随机森林模型对n个不同的决策树进行训练，最终预测时选取n个决策树结果中占比最大的预测结果。在训练过程中，随机选择样本点对每个决策树进行训练，并且在所有特征中随机选择m个特征作为决策树中节点分支的依据。我们通过对n个决策树的组合构建随机森林模型。在实验过程中为了探寻效果最好的模型，我们对不同的决策树个数、不同的决策树的最大深度都进行了训练，以得到精确度最高的随机森林模型。

5 训练和评估方法

本部分所叙述的训练和评估方法，指的是宏观上的指导原则。这些原则是在综合了第3部分和第4部分分析的基础上，针对本项目面临的人类行为分类任务所制定的，主要包含训练集、测试集和验证集的划分方式，以及评估模型的方法等。

5.1 划分训练集、测试集和验证集

在PAMAP2数据集中训练一个可供使用的人类行为分类模型是一个有监督模型训练任务。在有监督学习中，训练集和测试集的划分是一个必须考虑的问题。合理的训练集和测试集划分将有助于模型训练，而不合理的划分方法可能导致模型欠拟合或过拟合。

常用的训练集和测试集的划分方式之一是分层抽样。如果将其应用于本项目中，将针对每一种行为进行分层后抽样。然而，在本项目中，我们认为分层抽样的方式并不合适。原因有二：其一，本数据集是按照时间

顺序采样而得的，采样分层抽样的方式将极大程度上破坏这种时间顺序，可能不利于后续分析和处理；其二，我们希望训练出的模型在面对一个从未见过的佩戴IMU的志愿者的数据时也能做出正确的分类（即具有良好的泛化能力），而分层抽样产生的测试集可能无法有效地测试出这一能力。

因此，我们采用另一种划分方式，即抽取数据集中某几位志愿者的数据作为测试数据，其余志愿者的数据合并作为训练数据。根据统计，107号志愿者和108号志愿者的数据覆盖的活动行为类型最全面，且两者数据量之和占总数据的约24%，将这两位志愿者的数据合并作为测试集，其余数据合并作为训练集比较合适。

由于我们不仅希望找出最佳的模型，同时还希望在不同机器学习方法中调出合适的超参数，因此仅仅将数据集划分为训练集和测试集还不够。我们需要将训练集进一步划分，分出一部分作为验证集。于是，从总体上来看，我们将原始数据做如下划分：

表 1. 数据集划分方式

训练集	验证集	测试集
-----	-----	-----

划分验证集的方式依然可以使分层抽样，抽取25%的数据作为验证集，也可以是K-Fold方法，使用K折交叉验证，还可以是调出其中一位志愿者的数据作为Leave-one-out。在最终的版本中，我们分别尝试了这3种验证集划分方法。

5.2 模型评估指标

为了评估模型的能力，我们需要量化的指标。一个最自然的指标是模型的准确率(Accuracy)，即模型在测试集上的准确程度。然而，为了探索模型的稳定性，仅凭借准确率尚不足够。

在分类问题中，我们经常使用混淆矩阵(Confusion Matrix)来统计模型的预测结果。一个经典的混淆矩阵如下表：

表 2. 混淆矩阵

Confusion Matrix		预测值	
		正例	反例
真实值	正例	TP	FN
	反例	FP	TN

根据混淆矩阵，模型的准确率的定义为：

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

而对于分类问题而言，常用的评估指标还有：

✧ 精确率（又称查准率，Precision），即正确预测为正的占全部预测为正的比例。精确率代表对正样本结果中的预测准确程度，精确率的提出是让模型的现有预测结果尽可能不出错。

$$Precision = \frac{TP}{TP + FP}$$

✧ 召回率（又称查全率，Recall），即正确预测为正的占全部实际为正的比例。召回率是针对原样本而言的，其含义是在实际为正的样本中被预测为正样本的概率。高的召回率意味着可能会有更多的误检，但是会尽力找到每一个应该被找到的对象。

$$Recall = \frac{TP}{TP + FN}$$

✧ F1值（F1-score）。精确率和召回率互相影响，理想状态下肯定追求两个都高，但是实际情况是两者相互“制约”：追求精确率高，则召回率就低；追求召回率高，则通常会影响精确率。综合起来考虑它们，最常见的方法就是F1值，它是二者的调和平均。

$$F_1Score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}$$

通常而言，准确率越高，模型的分类效果越好；精确率越高，模型对负样本的区分能力越强；召回率越高，模型对真样本的识别能力越强；F1-score越高，说明模型越稳健。因此，在本项目中，我们将综合采用这4种评价指标，来判断模型训练的效果，选择最佳的模型。

需要指出的是，以上关于评价指标的讨论针对的是二分类问题，而我们的任务是一个多分类问题。在多分类问题中，我们统计的是每一类的4项指标：给定一类为正例，其余类均为反例，由此来计算这些评价指标，并最终对所有类别按照类别样例数目求取加权平均值。

6 训练过程和结果分析

基于第3部分对数据集的探索和认识及第4部分对各类模型原理的理解，在综合运用第5部分所述的训练和评估方法的基础上，本项目采取了逐点优化、逐层递进的方式，对各类模型进行训练和评估，在训练中总结优点和不足，并及时改进、开发新版本。在实验中，我们累计开发了4个大版本，其中包含7个小版本，以及附加的2个神经网络版本，共计9个版本的模型。下面将根据模型由简单到复杂的顺序详细介绍这9个版本。

6.1 版本1.0

在正式开始训练模型之前，我们希望找到一个评估模型的Baseline。因此，我们开发了1.0版本。这是一个极其原始的版本。在1.0版本中，我们仅按照第3部分所述对数据集中的缺失值加以填补，除此之外不做任何其余的处理。我们在此基础上，直接运行不含正则项的逻辑回归和普通决策树模型，对数据集进行学习。经过训练，模型的拟合效果如下表：

表 3. 版本 1.0

Model	Accuracy	Precision	Recall	F1-score	Training Time
逻辑回归	65.92	59.80	56.68	56.07	768.61
决策树	67.48	61.58	62.01	60.12	1622.45

（注：表中4项评估指标的值均为乘以100后的结果，训练时间的单位为秒，下同）

从训练结果可以看出，模型的准确性和稳定性均不高，且训练时间过长，时间成本过高。

6.2 版本2.0

在1.0版本的基础上，尝试对原数据集进行分段（分箱），并取每个特征在每段中平均值作为该段的特征（例如分段后的某一段心率为该段内原先所有心率值的平均值），使数据集大小由1942872个样本降低至3799个样本。样本数量的降低带来了训练成本的降低，同时可以让训练迭代更多的次数，提供准确率。经过测试，选取512个样本为一段比较合适，这可以使得所有不同周期的行为都至少涵盖2~3段，利于训练。

从此版本开始，后续所有版本都将尝试四类模型，分别是逻辑回归、K邻近、决策树（随机森林、AdaBoost）、支持向量机。在模型的超参数选择上，将依据模型在验证集上的表现选取合适的超参数，具体选择集合如下表：

表 4. 模型超参数选择范围

Model	Hyper-parameter
逻辑回归	正则化系数C~{0.001, 0.01, 0.1, 1, 10, 100, 1000, ∞}
KNN	欧拉距离/曼哈顿距离；K~[2, 19]
随机森林	树的最大深度~{10,20,30}；森林中决策树的数量~{50,120,200}
AdaBoost	基分类器循环次数~{5,10,25,50,75}；基分类器特征选择标准
SVM	C~{0.01, 0.1, 1, 10, 100}；核方法~{linear, rbf, sigmoid, poly}

版本2.0经过训练后的成绩表现如下表5：

表 5. 版本 2.0

Model	Accuracy	Precision	Recall	F1-score	Training Time
逻辑回归	62.05	70.39	54.44	54.71	7.49

KNN	69.90	76.76	60.63	60.45	56.56
随机森林	80.25	87.77	70.82	71.37	210.21
AdaBoost	71.14	62.71	63.50	61.87	33.81
SVM	61.94	71.04	55.31	56.07	158.75

由此可见，经过分段之后训练成本大大降低，所有模型的训练时间都在可以接受的范围之内。正是由于成本的降低，我们才能够迭代更多的次数、尝试更多的超参数组合。不过，该模型的准确率依然不是很高，只能达到和1.0版本相近的准确率。

6.3 版本2.1

2.0版本中，取每个特征在每段中平均值作为该段的特征。仅仅使用平均值的方式会损失每段内的很多信息，造成模型预测结果不准确。因此，我们考虑增加描述特征，来更好地描述每个特征在每段中的取值和变化情况。具体而言，除了使用平均值外，增加最大值、最小值、方差、偏度和峰度，即每一个特征现在使用6个值来进行描述。在此基础上训练的结果如下表6：

表 6. 版本 2.1

Model	Accuracy	Precision	Recall	F1-score	Training Time
逻辑回归	72.29	78.56	70.18	69.30	18.19
KNN	87.28	89.10	84.11	85.80	234.81
随机森林	90.17	92.89	88.73	89.98	408.29
AdaBoost	57.91	58.21	56.66	53.22	111.53
SVM	71.35	77.98	70.09	70.71	173.21

由此可见，经过增加描述特征后，模型的预测准确率有了显著的提升，模型的稳定性也有相应的提高。随机森林模型的准确率已经达到了90%。不过，高准确率也带来了训练时间的增长。

6.4 版本3.0

在版本2.1的基础上，我们希望对原始数据做进一步的处理和优化，以便更好地提取出其中的特征。根据第3部分的数据探索，我们注意到原数据集上可能含有冗余的信息。在每个动作开始前和结束后的一小段时间，传感器返回的数据对应的是静息状态下的数据，而非运动状态下的数据，因此为冗余数据。通过去除每个动作开始10秒和结束10秒的数据，对原数据集进行精简，去除冗余，而后再采用前文所述的分段方法，我们得到了3.0版本。其训练成果如下：

表 7. 版本 3.0

Model	Accuracy	Precision	Recall	F1-score	Training Time
逻辑回归	73.31	80.18	75.03	74.27	47.60
KNN	91.53	91.92	90.03	90.57	287.50
随机森林	92.34	95.55	93.59	93.56	238.67
AdaBoost	71.22	69.56	64.90	60.90	56.71
SVM	73.78	82.47	75.72	75.44	138.53

由此可见，精简数据集去除冗余后，各个模型的准确率均后不同程度的提高。

6.5 版本3.1和版本3.2

在版本3.0的基础上，我们进一步尝试了用第3部分中叙述的归一化方法，对数据进行处理。这两个版本分别使用数据标准化(Standard Scaler)和Yeo-Johnson Transform对原数据集进行归一化，并测试模型预测效果如下：

表 8. 版本 3.1： Standard Scaler

Model	Accuracy	Precision	Recall	F1-score	Training Time
逻辑回归	81.09	87.12	83.26	83.35	49.10
KNN	92.11	93.75	92.34	92.50	100.77
随机森林	92.34	95.54	93.59	93.56	246.82
AdaBoost	70.64	70.10	63.34	60.46	73.82
SVM	83.09	88.96	86.01	85.44	78.06

表 9. 版本 3.2： Yeo-Johnson Tranform

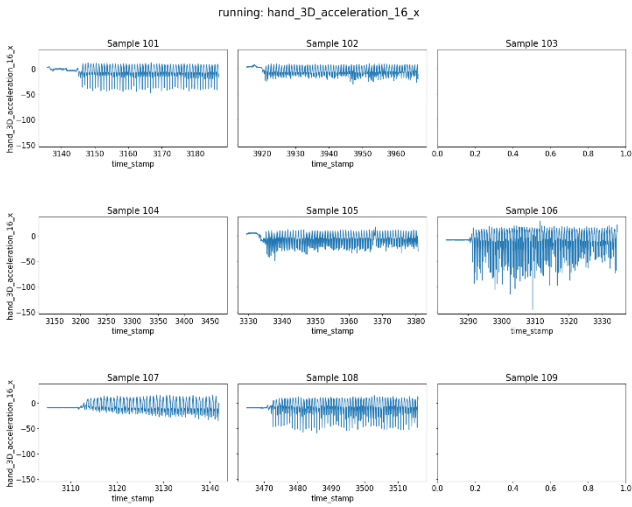
Model	Accuracy	Precision	Recall	F1-score	Training Time
逻辑回归	86.08	89.93	86.35	86.78	10.93
KNN	96.75	96.94	96.71	96.80	390.66
随机森林	92.22	95.32	93.36	93.42	201.66
AdaBoost	73.54	69.44	54.72	62.42	90.56
SVM	88.75	90.40	88.39	88.48	84.87

通过以上结果可见，使用归一化后，模型的预测效果都有了显著的提升，并且使用Yeo-Johnson Transform的效果要优于Standard Scale。可以猜测这一结果产生的原因，由于高斯分布特征最为显著，方差最为稳定，因此在多种机器学习模型中，高斯分布的数据分类效果最好，故Yeo-Johnson变换的效果最优。

6.6 版本4.0

在版本3.2的基础上，我们希望使用更高级的数据处理手段，进一步加强数据特征的提取。注意到，在很多活动行为中，传感器的数据随时间呈周期性变化，例

如，其中一名志愿者手部传感器在跑步时返回的数据如下图：



针对周期性时域数据，常见的处理方式是使用离散傅里叶变换，将其转换到频域中，得到每段数据的频谱结构和变化规律。如果将频谱信息加入训练特征中，将使得不同活动独特的特征更加清晰。将离散傅里叶变换后峰值最高的5个频率和振幅作为特征加入训练数据中，在新的数据集上训练，得到了4.0版本，效果如下表：

表 10. 版本 4.0

Model	Accuracy	Precision	Recall	F1-score	Training Time
逻辑回归	96.86	97.29	96.13	96.64	18.18
KNN	97.80	97.61	96.85	97.13	732.58
随机森林	98.14	97.73	96.52	96.87	176.32
AdaBoost	76.33	77.45	71.95	70.47	91.50
SVM	97.68	97.66	96.76	97.14	324.23

由上表可见，加入了频域信息后，所有模型的准确率均有所提升，除AdaBoost外其余模型的准确率均达到了95%以上，随机森林模型的准确率最高，达到了98%。因此，我们将该版本作为最终版本，认为在这样的处理模式下模型良好地实现了人类活动行为分类的任务。

6.7 神经网络版本

在版本3.0和版本4.0的数据处理方式的基础上，我们还尝试了使用神经网络(多层感知机)技术进行分类。我们使用的神经网络模型包含两个隐藏层，分别含有4096和2048个神经元，并使用ReLU函数作为激活函数。输出层采用Softmax函数，将数据映射为12个类别的概率，并选择概率最高的一类作为预测结果。在两个隐藏层之后，我们添加了概率为0.5的Dropout层来防止过拟合的产生。两个神经网络版本的测试效果如下：

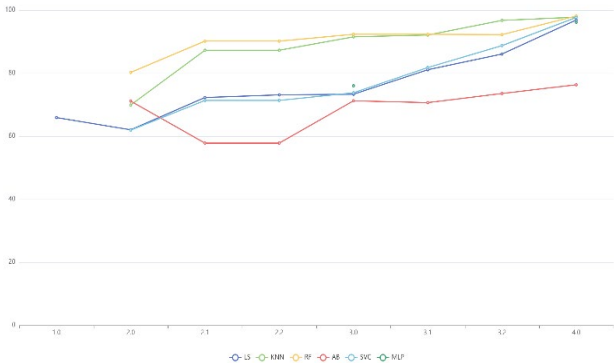
表 11. 神经网络版本

Model	Accuracy	Precision	Recall	F1-score	Training Time
基于3.0版	76.01	75.98	73.82	73.42	372.08
基于4.0版	96.17	96.91	94.93	95.73	387.29

可以看出，神经网络和其他模型所达到的正确率和稳定性相仿，在3.0版本的基础上神经网络可以达到76%的准确率，而在4.0版本的基础上可以达到96%的准确率。

7 结论和启示

在综合了4个大版本、9个小版本的实验结果后，绘制准确率曲线如下图：



由该曲线可以看出，每次版本迭代后，各个模型的准确率均有不同程度的提升，最终逻辑回归、KNN、随机森林、SVM、神经网络都达到了95%以上的准确率。其中，逻辑回归训练时间最短；随机森林准确率最高；SVM的F1-score最高，模型最稳定。

从模型训练的过程来看，数据的处理至关重要。良好的数据处理范式能够更有效地提取出数据中的特征，从而能够使模型拥有更好的准确度和稳定性。

利用训练集-验证集-测试集的训练范式，我们同时还得到了各个模型最优的超参数，如下表：

表 12. 模型最优的超参数选择

Model	Hyper-parameter
逻辑回归	正则化系数C = ∞ （不使用正则化项）
KNN	曼哈顿距离；K = 4
随机森林	树的最大深度 = 20；森林中决策树的数量 = 200
AdaBoost	基分类器循环次数 = 50；基分类器特征选择标准 = 'entropy'
SVM	C = 10；核方法 = 'rbf'（高斯核）

不过，从实验结果来看，数据处理所带来的模型效果的提升程度远大于调整合适超参数带来的效果提升。因此，数据处理比模型原则以及超参数调整更为关键。

综合准确率、稳定性和训练时间成本来看，在人类活动行为分类任务中，选取随机森林模型、支持向量机模型或逻辑回归模型最为合适。

8 附录

本项目所使用的的数据集PAMAP2下载网址如下：
<https://archive.ics.uci.edu/ml/datasets/PAMAP2+Physical+Activity+Monitoring>

本项目的相关代码使用Python3编写，代码开源网址如下：
<https://github.com/yyChen233/Human-Activity-Classification>

参考文献

[1] B; Aarts R.M Long X.; Yin. “Single-accelerometer based daily physical activity classification”. In: Proceedings of the EMBC. IEEE, 2009.

[2] Panu Korpipa Juha Parkka Miikka Ermes. “Activity Classification Using Realistic Data From Wearable Sensors”. In: IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE. IEEE, 2006.

[3] Youngwook Kim and Hao Ling. “Human Activity Classification Based on Micro-Doppler Signatures Using a Support Vector Machine”. In: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. IEEE, 2009.

[4] Thomas Plotz Nils Y. Hammerla, Shane Halloran. “Deep, Convolutional, and Recurrent Models for Human Activity Recognition using Wearables”. In: Open Lab, School of Computing Science, Newcastle University, UK, 2016

[5] Attila Reiss. PAMAP2 Dataset: Physical Activity Monitoring.

[6] Li Dou. Machine Learning for Predictive Data Analytics Course PPT