



汕头大学
SHANTOU UNIVERSITY

毕业论文（设计）

题 目： 基于人体关节夹角的人体动作识别算法

学 院： 工学院

系 别： 计算机系

专业年级： 19 大数据

学生姓名： 吴逸南

学 号： 2019612039

指导教师： 赵志丹

完成时间：2023 年 5 月

基于人体关节夹角的人体动作识别算法

摘 要

人体动作识别是当今一个十分热门的话题，而且人体动作识别也是渗透到了当今的信息科技的许多方面例如视频检索、身份识别等。人体动作识别一直是计算机视觉领域里面的一个十分含有技术含量的课题，而人体动作分类又是人体动作识别中的一项比较关键的组成部分。

随着现代机器学习方法技术的发展，对于这个问题越来越多的相关的研究方法被提出。但是这些方法都具有一定的缺点：基于手工提取特征的方法往往会对会受到一些噪声的影响，而基于深度学习的算法又会耗费许多的算力。对此，我们提出了一个基于人体关节夹角的人体动作序列特征抽取方法。为了更好地进行分类任务，我在论文当中使用的分类模型是一个基于软投票策略的 Bagging 的集成学习分类模型。

论文中提出的方法结合了传统机器学习方法和深度学习方法的优点：先用较为成熟的深度学习方法的预训练模型先从噪声比较多的人体动作数据之中把人体动作的关键点信息提取出来，这样就使得算法的结果不会被一些噪声给影响到。而后再使用可解释性强，耗损的算力低的传统的手工设计特征加上机器学习的方法进行建模。

我们先使用这个人体动作序列特征抽取方法抽取出特征向量，然后把抽取出的特征向量进行特征选择。而后使用进行了特征选择以后的特征向量进行模型训练以及模型验证且训练集和验证集的比例划分为 5:5。在本论文当中使用了 G3D 数据集作为数据集，最终本论文提出的方法在 G3D 上的分类准确率为 92.3%。该结果验证论文中方法的可行性。

关键词：G3D，人体关节夹角，Openpose，集成学习

Human Action Recognition Algorithm via Human Joint Angle

Abstract

Human motion recognition is a very hot topic today, and human motion recognition has penetrated into many aspects of information technology today, such as video retrieval, identity recognition, etc. Human action recognition has always been a very technical subject in the field of computer vision, and human action classification is a important component of human action recognition.

With the development of machine learning, more and more related methods have been proposed. However, the shortcomings of this kind of method is obvious: the method based on manual feature extraction is often affected by some noise, And the method based on deep learning will consume a lot of computing resource. Hence, we propose a human action sequence feature extraction method based on the angle between human joints. In order to accomplish the classification task better, the classification model that we used in the thesis is an ensemble learning classification model based on Bagging strategy of soft voting .

The method proposed in the paper combines the advantages of traditional machine learning methods and deep learning methods: we firstly use the pre-training model of deep learning method to extract the keypoint of human motion from the human motion data . Therefore , the result of the method will not be affected by some noise. Then we use traditional manual design features with strong interpretability and low computational consumption and machine learning methods for modeling.

We first use this human action sequence feature extraction method to extract feature vectors, and then apply feature selection to extracted feature vectors. Then use the feature vector which go through feature selection for model training and model verification, and the ratio of training set and verification set is divided into 5:5. In this paper, the G3D data set is used as the dataset, and finally the method proposed in this paper has a performance of 92.3% on G3D. This result verifies the feasibility of the method in the paper

Keywords: G3D, Human Joint Angle, Openpose, Ensemble Learning

目 录

1 概 述	1
1.1 人体动作识别领域	1
1.1.1 人体动作识别的背景	1
1.1.2 人体动作识别的方法分类	1
1.2 人体动作分类	2
1.2.1 直接分类	3
1.2.2 时序方法	3
1.2.3 时空方法	4
1.2.4 基于人体部位的方法	4
1.2.5 基于 Manifold 的学习方法	5
1.2.6 中等特征学习方法	5
2 特征工程	7
2.1 人体关节坐标的获得	7
2.2 人体动作特征向量的获取	9
2.3 特征选择	11
3 机器学习建模	13
3.1 所用到的分类模型	13
3.1.1 logistic 回归	13
3.1.2 支持向量机	14
3.1.3 KNN	15
3.1.4 基于树的分类算法	16
3.2 集成学习	17
3.2.1 Bagging 算法	18
3.2.2 Boosting 算法	18
3.2.3 我们的机器学习模型	19
4 在 G3D 数据集上的实验	21
4.1 G3D 数据集	21
4.2 Openpose 获取关键点的 2d 坐标	22
4.3 3D 坐标的引入以及动作特征向量的提取	23
4.4 特征选择	24
4.5 机器学习分类模型调参训练	24
4.5.1 logistic 回归	25
4.5.2 支持向量机	25
4.5.3 KNN	26
4.5.4 随机森林	27
4.5.5 Xgboost 和 lightgbm	28
4.5.6 集成学习模型的分类效果	28
5. 结论	30
致谢	32
参考文献	33

1 概 述

1.1 人体动作识别领域

人体动作识别是当今计算机视觉领域一个非常热门的话题。这个领域所涉及到的知识以及相关的研究课题是非常多的。下面我对人体动作识别的背景以及人体动作识别的方法做一个简单的介绍。

1.1.1 人体动作识别的背景

人体动作识别是当今一个十分热门的话题，并且人体动作识别也是渗透到了当今的信息科技的许多方面，如智能视频监控和环境家庭监控、视频检索、身份识别等^[1]。对人体的动作行为进行识别这个总的大课题也涉及到了 CV 领域的一些相关的课题，包括人体检测，人体姿态估计，人体动作跟踪等等。因为人体动作行为的完成过程是一个连续的过程，所以在研究人体动作行为的过程中需要用到时间序列建模的相关知识。因为遇到的困难比较多，涉及到的方面也比较广，所以人体动作识别一直是计算机视觉领域里面一个十分具有挑战性的课题

如何良好地对人体的动作进行建模和如何找到一种可以抽取鲁棒性很强的特征表示的特征提取方法是人体动作识别中十分重要的工作。和图像的特征提取相比，对视频当中的人体动作的特征抽取是在其基础上的一个拓展，视频中的人体动作不仅仅表现了人体动作的外观，也表现了这些外观之间随着时间的变化，所以相比起图形空间的特征提取，人体动作的特征提取增加了“时间”这一维度。也就是说特征表示从二维提升到了三维。近年来，研究者们提出了许多人体动作的特征抽取方法，包括基于时空变化的局部特征和全局特征^[2]、基于关键点跟踪的轨迹特征^[3]、基于运动变化的运动特征^[4]，基于深度信息^[5]，以及基于人体姿态的动作特征变化^[6]。自从深度学习的出现并在非结构化数据的机器学习任务中取得非常好的效果以后，各种各样的具有十分强大的拟合能力的深度学习模型被提出。这些模型可以自发地从视频当中提取出优秀的人体动作行为的特征表示。

从数据的角度来看，大多数关于人类行动识别的数据集可以划分为一下三种：**RGB** 数据、深度数据或骨骼数据。随着深度摄像机的发展，深度数据在人体行为识别所占据的作用是越来越显著了，这种作用尤其体现在通过人体姿态估计来提取人体骨骼数据。此外，对 **RGB** 视频中对人体进行检测以及对视频中的姿态进行估计的^[7-9]的发表的相关论文表明，具备强大拟合能力的深度学习模型可以在场景极其复杂的环境下对多人姿态估计这一任务可以取得非常不错的效果。这些年来也在不断地发表基于深度序列和骨骼数据的人体动作识别的相关的论文^[10-11]。这些发表的论文里的方法解决了对于 **RGB** 摄像机中如何对人类的行为进行识别的一些关键性问题，而且识别的效果也是相当不错。

1.1.2 人体动作识别的方法分类

从研究的数据集的形式来划分的话，人类动作识别的研究方法可以大致分为以下两种：

基于 RGB 数据的方法，基于 RGB 数据和深度数据相结合的方法的方法。随着机器学习技术的蓬勃发展，在这些数据上的基本研究方法可以大体划分为以下两种方法：一种是基于手工设计进行传统特征提取的方法。另一种方法就是基于深度学习的方法。深度学习具有令人难以想象的拟合能力，而且深度学习方法可以实现端到端的机器学习任务。无论是什么样的数据类型以及采取什么样的技术进行计算，最关键的都是提取合适并且具有非常高的鲁棒性的人体动作的特征表示。

近几年来科学家们基于 RGB 数据提出了许多种类型的特征抽取方法，如基于时空体积的特征^[12-13]、时空感兴趣点特征^[14-15]和联合轨迹特征^[16-17]。但是，因为现实生活以及研究之中总是会遇到一些有干扰的东西例如摄像机运动、遮挡、复杂场景以及人工检测和姿态估计方法的局限性，所以基于手工提取特征的那一套传统机器学习路子就受到了影响，所以如何设计出一套可以充分忽略这些影响的具有十分强的鲁棒性的手工特征提取方法是走这一套路子的关键。

因为不论环境以及背景如何地变化，视频或者图片的深度数据基本都是稳定不变的，所以基于深度传感器提供的数据可以实现实时、鲁棒性十分强的人体姿态估计。基于深度信息和骨骼序列的人类动作识别方法可以实现非常高的人体动作识别效果，而且所需要的时间成本以及计算算力都不是很高。所以这种人体动作的识别流派受到了很多研究者的热烈欢迎，并且也有不少的研究成果。但是这种方法也是具有局限性的，主要的局限性体现在深度传感器的获取图片深度信息的准确性。有三种常见类型的常用深度相机：三角测量，飞行时间相机，和基于结构光的相机。结构光和基于 TOF 的深度传感器对光的强度非常敏感，所以很容易受到影响，这样会造成深度数据的准确率以及人体动作识别的效果都会大打折扣。三角测量系统的成本较低，但是要综合利用三角测量的信息从而计算出真实的深度数据，这就使得深度信息计算的复杂性较高，而且在环境比较暗的情况下这种方法的效果非常低。激光扫描仪也可以用于深度信息的获取，但是这些设备的价格太过昂贵。

与走传统机器学习路线的手工设计抽取方法进行抽取的人体动作特征表现方法不同，深度学习方法在进行端到端学习的过程当中可以自动学习人体动作的特征，并且特征效果提取是非常良好的。毫无疑问这对人体动作识别领域的研究又打开了一扇新的门。人体动作识别可以利用到的数据以及有效的特征是非常多的，例如外观/图像信息、光流序列、深度序列和骨架序列，深度网络的多模态特征学习可以同时把这些数据以及特征都给利用起来，从而得到更加准确，更加具有鲁棒性的人体动作的特征表示。

而人体动作分类又是人体动作识别中的一项比较关键的组成部分。在目前的研究中，只有动作分类问题涵盖了到了人体动作识别领域几乎所有的方面。

1.2 人体动作分类

对人体动作进行特征抽取从而得到对应的人体动作向量以外。接下来就可以以这些特征作为训练的数据，应用各种各样的机器学习知识来训练对应的动作分类器来进行动作的分类。动作分类方法可以分为以下几种类型：

1.2.1 直接分类

这样的方法就是把一个动作序列先变成一个特征向量，然后使用现成的分类器直接识别动作，如支持向量机^[18]，k-最近邻^[19]等。在这些方法中，人体动作序列的动态模式是通过人体动作在执行过程当中的人体形状以及通过词袋模型（它使用视觉单词的直方图来编码局部运动模式的分布）这两种方法来进行体现的。

单词袋的方法在再过去一段时间里引起了许多研究者的研究兴趣。如图 1-1 所示，这些方法首先使用时空兴趣点检测器^[20]来检测局部突出区域。而后在三维的时空兴趣点那里进行一些传统的人体动作特征提取例如光流以及梯度，但是这样提出到的特征有时候会显得比较冗余，所以接下来就要采用主成分分析的方法使特征降维。然后，可以通过 k-means^[21]或 Fisher vector^[22]来计算视觉单词。最后，经过这一系列流程的特征抽取，一个复杂的人体动作行为序列就可以转化为视觉单词的直方图，这意味着“人体动作”这种非结构化数据转化成了传统的向量，而后就可以使用经典的机器学习分类器例如支持向量机，KNN，逻辑斯特回归进行分类。在之后的研究当中，词袋模型逐渐体现了它的一个优秀的特性：它对人类的外观以及人体的姿态变化的不敏感性。然而，他们忽略了两个问题：第一是人类的行为是一种在时间上的连续行为，是具有时空性的。第二是人类的外观并不是一些互相独立的部件构成，人体每一个部件之间都是具有联系的，也就是说人类的姿态具有结构性。这两个问题可以通过序列的方法^[23]和时空方法^[24]来解决。



图 1-1：词袋模型建模方法的流程图

1.2.2 时序方法

生活中以及研究当中我们很多的数据都不仅仅只是连续的而是相互之间都是有联系的，例如语音，音乐，文本，动作。在研究当中序列的特点在于某一时刻的数据不仅和现在有关，而且和过去的数据以及未来的数据也具有很大的关系。因此针对这种类型数据的建模方法是一定要把各个阶段的数据之间的连续关系发现并且利用这些联系关系进行建模，从而得到更加丰富，更加全面的建模结果。人体动作显然就是那种满足时序性的非结构化数据，因此可以使用序列状态模型，如隐马尔可夫模型^[25]，条件随机场^[26]和结构化支

持向量机^[27]。这些方法是把一个动作视频看成一个一个连续的动作视频帧组成的时间序列模型。在^[25]的研究过程当中研究人员对一个房间中的人类日常轨迹进行研究，并且使用两层的隐马尔可夫模型来对房间里的人类行为轨迹进行建模。在^[27]的研究表明，如果要更好，更合理地对人体动作行为进行建模，我们可以选择对人体动作里面比较关键的那一部分姿态进行建模。实验也证明了这样可以取得不错的效果。究其原因就是因为这样的方法摒弃了人体动作之中一些十分冗余，对于动作分类器的训练没有研究价值的动作信息，某种程度上这也算是一种特征选择。这样的方法都是通过把一个视频分解成一个一个独立的视频帧，而后从这些视频帧当中提取整体的特征。但是这些特征具有一个很显著的缺点，就是这些特征背景噪声很敏感，而且通常在那些因为背景噪声难以剔除而比较具有挑战性的数据集上，通过这种方法所提取出来的。特征往往不具有比较良好的效果。

1.2.3 时空方法

虽然 1.2.1 的方法在一些动作数据集上显示出了良好的结果，但这些方法没有考虑局部特征之间的时空相关性，在时空兴趣点当中的一些隐藏的具有研究价值的信息。这个问题在^[28]中得到了解决，它使用相对坐标特征学习全局高斯混合模型，并使用多个高斯混合模型来描述兴趣点在局部区域的分布。在^[29]中提出了兴趣点上的全局特征来捕获兴趣点的详细几何分布。该特征是通过使用定义为扩展的三维离散氦变换来计算的。该特征捕获感兴趣点的几何信息，对几何变换保持不变，对噪声保持鲁棒性。感兴趣点的时空分布是由^[30]中的方向性金字塔共现矩阵来描述。DPCM 表征了局部特征的共现统计量以及并发特征之间的时空位置关系。图是对结构化对象建模的强大工具，在^[31]中用于捕获局部特征之间的空间和时间关系。以局部特征作为双图模型的顶点，并将帧内和帧间的局部特征之间的关系用边缘来表征。在^[31]中提出了一种新的上下文相关图核族来度量双图模型之间的相似性。虽然上述方法已经取得了有希望的结果，但它们仅限于小数据集，因为它们需要建模兴趣点之间的相关性，这在大数据集上非常难以做到的，因为时间复杂度太高了。

1.2.4 基于人体部位的方法

人体是一个结构化就是说人体之间的各个部位的运动并不是互相独立而是相互联系的。所以通过研究人体各个部位来研究人体的运动情况是一个非常好的方法。在^[32]中提出了一个星座模型，该模型模拟了身体部位的位置、外观和速度。受^[32]的启发，在^[33]中提出了一个基于部分的层次模型，其中部分由模型假设生成，局部视觉词由身体部分生成。^[34]中的方法将局部视觉词视为零件，并对零件之间的结构信息进行建模。这项工作^[35]中得到了进一步的扩展，作者假设一个动作是由多项分布生成的，然后每个视觉单词都是由基于动作的分布生成的。这些基于部分的生成模型被判别模型进一步改进。在^[36]中，某一部分被认为是其模型中的一个隐藏变量。它对应于一个正能量最大的显著区域。

1.2.5 基于 Manifold 的学习方法

人类动作视频可以用时间上变化的人体轮廓来描述。然而，这些剪影的表现通常是高维的，并阻碍了我们有效的动作识别。为了解决这一问题，在^[37]中提出了流形学习方法来降低轮廓表示的维数，并将其嵌入到非线性的低维动态形状流形上。^[38]中的方法采用核主成分分析进行降维，发现了流形中动作的非线性结构。然后，利用双链分解的 CRF 模型将低维空间中的剪影特征分类为人类行为。在^[39]中提出了一种新的流形嵌入方法，该方法可以使与不同类的轮廓相关联的时间子空间之间的主角度最大化。这个方法往往会取得非常好的效果，但是这种方法的一个基本要求是获得一份非常优秀，没有缺陷的人体轮廓数据，这往往在日常生活当中是做不到的。

1.2.6 中等特征学习方法

词袋模型已经显示出对背景噪声的鲁棒性，但可能不足以表达来描述存在较大的外观和姿态变化时的动作。此外，由于低级特征和高级动作之间的语义差距很大，它们可能不能很好地表示动作。为了解决这两个问题，提出了层次化方法^[40]来学习额外的表示层，并期望更好地抽象低级特征进行分类

层次化方法从低级特征中学习中层特征，然后用于识别任务。学习到的中级特征可以被认为是从用于培训的同一数据库中发现的知识或由专家指定的知识。最近，语义描述或在动作识别中得到了广泛的研究。这些语义被定义并进一步引入到活动分类器中，以表征复杂的人类行为^[41]。其他层次方法，如^[42]从观察的帧中选择关键姿态，也在模型学习过程中学习更好的动作表征。由于使用了人类的知识，这些方法显示出了优越的结果。

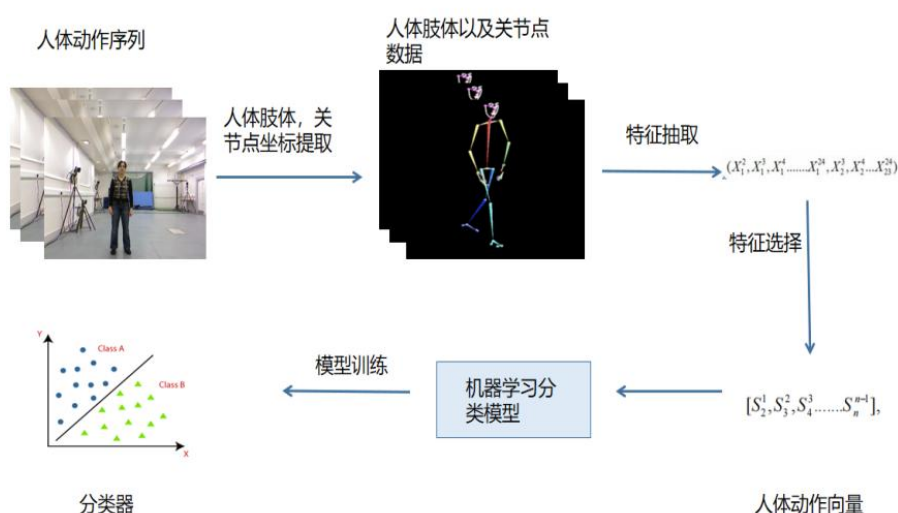


图 1-2: 人体动作识别算法的流程

通过对以往的人体识别方法以及对人体分类算法的回顾，我们可以将其分为两种类型：基于人工设计的特征提取的分类方法以及深度学习方法。但是基于人工设计的特征提取的

方法很容易受到噪声的影响，基于深度学习的方法虽然比较适用于有噪声的环境当中，但是其可解释性比较差，而且需要的算力比较多。为了克服这两个问题，我们提出了一种结合了这两种方法的优点的方法：先用较为成熟的深度学习方法的预训练模型先从噪声比较多的人体动作数据之中把人体动作的关键点信息提取出来从而进一步提出人体关节夹角的信息，而后再使用传统的手工设计特征加上机器学习的方法进行建模。这种方法既可以解决背景噪声比较多的问题，又具备良好的可解释性。上面图 1-2 是则是基于人体关节夹角的人体动作识别算法的流程图。

在接下来的文章中，我们在第二节介绍了特征工程的详细过程，而后在第三节之中说明了我们所提出的集成学习机器学习分类模型。分类模型的实验结果将在第四节详细给出。并且我们在第五节对第四节的实验结果进行了详细的讨论以及从实验结果当中也分析出了对该方法未来的展望。

2 特征工程

特征工程或特征提取或特征发现是利用领域知识从原始数据中提取特征的过程。其动机是利用这些额外的特征来提高机器学习过程的结果的质量，而不是只提供原始数据给机器学习过程。在本论文当中特征工程分为了一下三个部分：人体坐标的获取，人体特征向量的提取，特征选择。

2.1. 人体关节点坐标的获得

我们的特征抽取方法的第一步就是获取人体的关键点坐标。人体关键点检测是人体动作识别里面一项比较基础的任务。在我们的算法当中也要先执行这一步，为了获取人体关键点信息，我们运用了 Openpose^[9]这一人体关键点识别库，这个识别库可以比较准确地提取出人体关键点的坐标和人体的肢体表示，而且提取的结果对视频里面的噪声不敏感，具有比较良好的鲁棒性。

Openpose 的核心技术就是采用 PAF 进行自下而上的人体姿态估计，可以巧妙地解决多人肢干连接问题。该技术借鉴 CPM^[43]的方法，CPM 先对输入进行处理，最后得到的检测结果是一幅预测人体关键点的 heatmap。这样每一个关键点在 heatmap 上面都服从一个高斯分布，而后高斯分布的极大值点的位置就是对应的关键点的位置。

Openpose 的改进就在于它引入了 PAF 的概念。PAF 就是人体动作的亲合力场，通俗地说就是 PAF 代表了两个关节点属于同一个肢体的可能性，PAF 越大两个关键点之间属于同一个肢体的可能性越高。 F 是输入的图片通过 VGG-19 网络前 10 层经过预处理之后的结果。而后 F 作为输入兵分两路进入两个卷积神经网络进行学习，分别用以对人体的关键点和 PAF 进行回归预测。Openpose 的总网络结构图如下图所示：

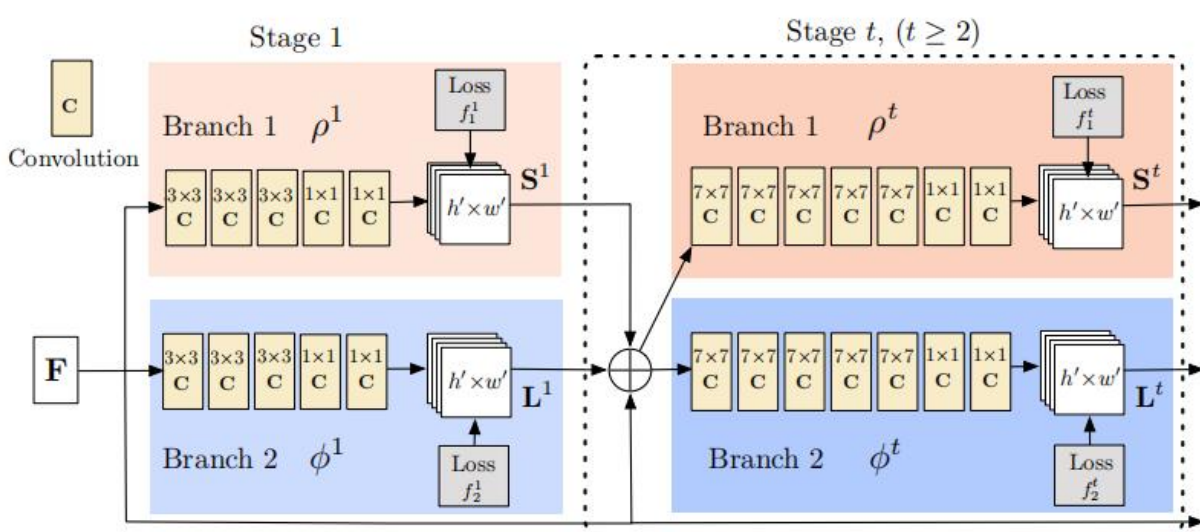


图 2-1: Openpose 的网络结构图（引自[9]）

F 经过这两个网络结构之后可以得到两个处理结果：一个是人体关键点的 heatmap，一个是 PAF 的 heatmap。而后把任意两个关节点（由 CPM 的网络结构预测出来的）之间的 PAF 进行线积分，就可以得到这两个关节点之间的 PAF。通过以上方法求出来的人体关键点以及对应的 PAF 可以按照图论的角度进行建模，这样接下来就把人体关键点的识别问题转化成了多个二分匹配的问题（如下图所示），然后我们可以使用图论的二分匹配算法例如匈牙利算法[47]得到最终结果。

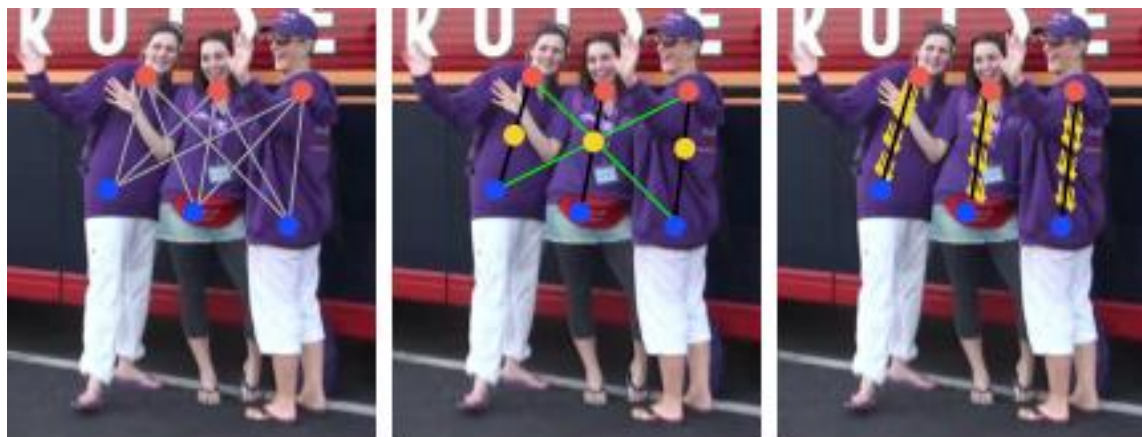


图 2-2: Openpose 的二分匹配问题实例（引自[9]）

MPII 数据集是一个用于衡量人体姿态识别效果的一个很好的数据集。数据集由将近 25000 张图片组成。数据集里面的图片是来源于人类每天的活动，并且还对人类每天的活动进行了分类。这个数据集里面有 400 多种人类活动的种类，并且每一个种类都有着对应的标签。Openpose 在 MPII 上面的表现如下图所示：

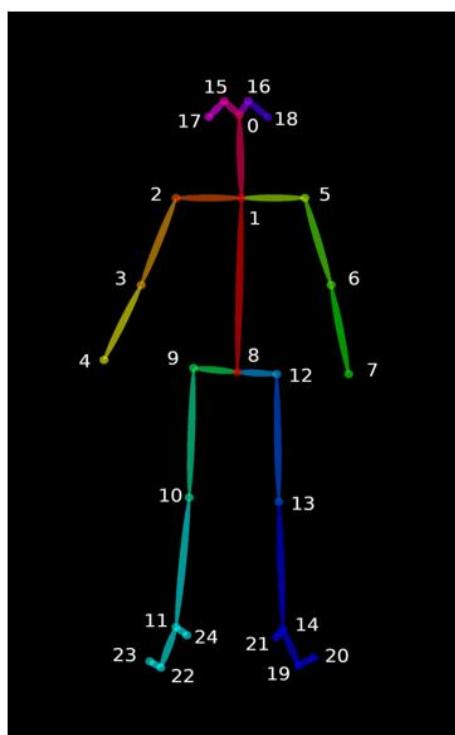
Method	Hea	Sho	Elb	Wri	Hip	Kne	Ank	mAP	s/image
Subset of 288 images									
Deepcut	73.4	71.8	57.9	39.9	56.7	44.0	32.0	54.1	57995
Iqbal et al.	70.0	65.2	56.4	46.1	52.7	47.9	44.5	54.7	10
DeeperCut	87.9	84.0	71.9	63.9	68.8	63.8	58.1	71.2	230
Ours	93.7	91.4	81.4	72.5	77.7	73.0	68.1	79.7	0.005
Full testing set									
DeeperCut	78.4	72.5	60.2	51.0	57.2	52.0	45.4	59.5	485
Iqbal et al.	58.4	53.9	44.5	35.0	42.2	36.7	31.1	43.1	10
Ours (one scale)	89.0	84.9	74.9	64.2	71.0	65.6	58.1	72.5	0.005
Ours	91.2	87.6	77.7	66.8	75.4	68.9	61.7	75.6	0.005

表 2-1: Openpose 在 MPII 数据集上面的表现（引自[9]）

所以 Openpose 可以提取出鲁棒性极强，不怎么受环境干扰的关键点数据。这比较好解决了噪声对算法执行效果的影响。

2.2 人体动作特征向量的获取

Openpose 是以 caffe 为框架开发的开源库，在这里我们直接使用了论文已经完成训练的预训练模型进行关键点抽取。训练出来的以训练模型有两种：body25 和 body17。这两种模型的区别在于一个可以识别出 25 个关节点一个可以识别出 17 个关键点。为了更加详细的描述人体的姿态，我们选择了 body25 作为预训练模型，body25 预训练模型所提取出来的人体骨架以及对应的人体关键点如下所示：



{0, "Nose" },	{12, "LHip" },
{1, "Neck" },	{13, "LKnee" },
{2, "RShoulder" },	{14, "LAnkle" },
{3, "RElbow" },	{15, "REye" },
{4, "RWrist" },	{16, "LEye" },
{5, "LShoulder" },	{17, "REar" },
{6, "LElbow" },	{18, "LEar" },
{7, "LWrist" },	{19, "LBigToe" },
{8, "MidHip" },	{20, "LSmallToe" },
{9, "RHip" },	{21, "LHeel" },
{10, "RKnee" },	{22, "RBigToe" },
{11, "RAnkle" },	{23, "RSmallToe" },
	{24, "RHeel" },

图 2-3: Body25 的人体骨架识别图（引自[43]）

一些传统的人体特征表示设计简单，需要的计算的算力虽然不多，时间复杂度也不高，但是特征提取的效果却不是那么的好。而使用深度学习方法训练所需要的数据集太大，耗费的计算资源太多，且训练的时间也太长。所谓我们需要找出一种鲁棒性比较强的动作特征表示方法。基于人体关节夹角的方法就是一类非常不错的选择。首先人体的关节夹角是人体骨架的一个非常明显且重要的特征，所以它可以比较准确地对人体骨架进行建模。并且人体关节点不会受到太多背景噪声例如光线的影响。此前，在《基于 OpenPose 的动作序列相似度计算算法》^[43]里面坐着这样的人体动作特征提取方法。在该论文里面的相关实验也证明了这种方法的可行性。该特征提取的方法如下所示：

①采样：假设要输入的动作序列为 S。我们可以将一个动作序列看做是一个图片的序列集合。所以我们采用滑动窗口采样法作为我们的采样方法，具体的方法就是等间距取 x 张图片，然后把这 x 张图片作为视频的代表。在^[43]实验中 x 取 50 的时候效果最好，所以论文当中 x 值也取 50。采样过程如下图所示：

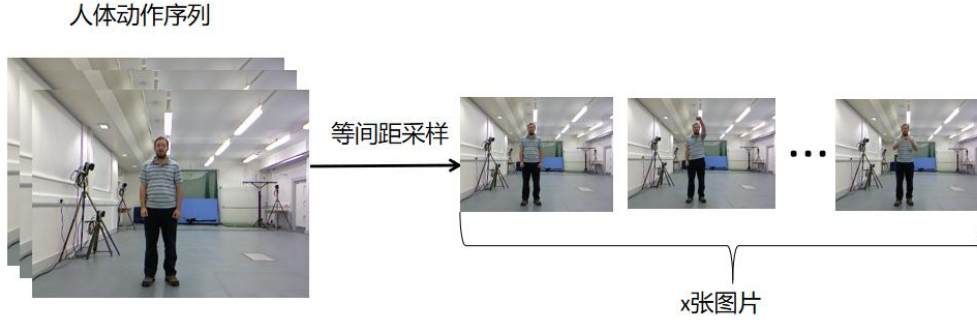


图 2-4: 等间距采样流程图

②单张图片的特征抽取: 基于 OpenPose 的 body25 预训练模型, 我们可以得到 25 个人体关键点坐标和 24 个人体肢体。而后我们可以根据以下的公式提取人体肢体向量:

$$\text{limb}_{ij} = \begin{cases} (x_i - x_j, y_i - y_j, z_i - z_j,) & \text{关节点}i \text{与关节点}j \text{之间存在肢体} \\ 0, & \text{关节点}i \text{与关节点}j \text{之间不存在肢体} \end{cases} \quad (1)$$

依照上面的公式, 每一张图片我们都可以得到 24 个非零肢体肢体向量。然后通过计算 24 个非零肢体向量两两之间的夹角的余弦值, 可以得到 276 个夹角余弦值。这样一张图片就转化成了一个 276 维的向量。这就是我们算法的单张图片特征抽取, 流程如图 7 所示:

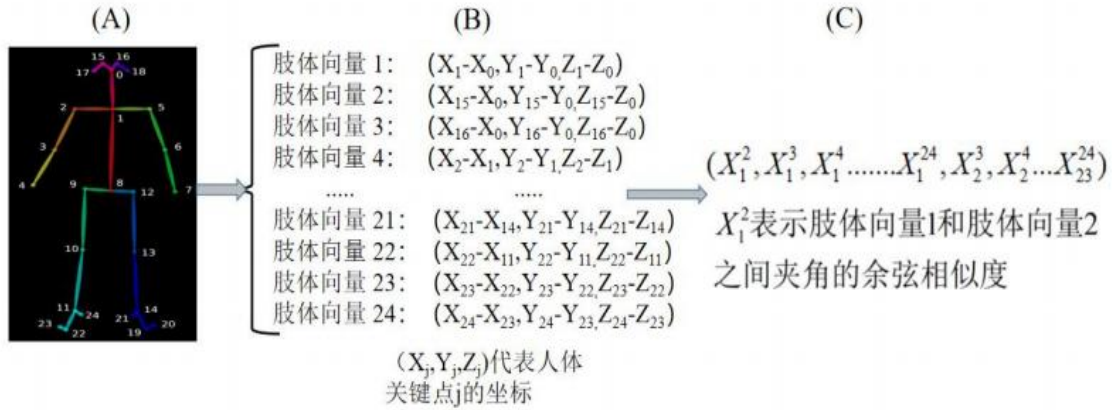


图 2-5 单张图片动作特征向量的提取 (引自[9])

③将 50 张图片都进行一遍特征提取, 那么就得到了 50 个动作特征向量, 然后把这 50 个动作特征向量按照顺序依次拼接成一个向量, 这个就是我们要求的动作特征表示。对单张图片的特征提取考虑了人体动作的结构性, 等距离采样的方法以及拼接特征向量的做法考虑了动作的时序性。所以这个人体动作特征表示可以比较有效地对人体动作进行建模。

2.3 特征选择

现在信息科技高速发展，获取数据的能力也随着不断提高。现在数据的处理已经延伸到了对于各种各样的高维，非结构化的数据处理。数据的快速增长无疑地该数据处理以及数据分析带来了比较高的难度。所以在数据的处理过程当中数据分析师们会经常使用一些数据挖掘以及机器学习的相关知识来处理数据，并且在数据当中发现数据所隐藏的知识。

当使用数据挖掘和机器学习算法来处理那些维度非常高的数据时，会出现一格非常严重的问题叫做数据的“维数诅咒”。因为维数过高，使得数据点在这些高维空间里面的距离过大，这样就会造成数据之间十分稀疏。这样不仅仅隐藏了数据本身所具有的知识，而且因为传统机器学习的算法都适用于低维数据而造成机器学习算法的效果会事倍功半。由于特征实在太多，往往会导致学习机器学习模型会在学习的过程当中发生过拟合的现象。这样会使得训练出来的模型在实际的应用当中表现也会十分的不好。从数据的存储角度来看，高维数据的存储也会耗损许许多多的存储资源。

数据降维是一个解决上面由于高维数据所导致的问题的一个非常有效的方法。它主要分为特征提取和特征选择两个主要组成部分。特征提取将原始的高维特征通过某种操作（在这里可以理解为一种映射）变成一个低维的特征向量表示。新构造的特征空间通常是原始特征的线性或非线性组合。特征选择是直接选择相关特征的一个子集来构建模型。

现实数据一般都会有这许多的特征，但是并不是所有的特征是可以为数据挖掘及机器学习算法提供有用的信息的。例如，在图 2-8 中，特征 f_1 是有价值的特征，而在 f_1 存在的基础之上 f_2 和 f_3 是多余的，因为它们没有能力可以对分类的工作提供任何有用的信息。因此，去除 f_2 和 f_3 不会降低分类模型的性能，反而可以节省计算资源，用于存储更多有用的特征，这样还能有可能提高模型的表现。

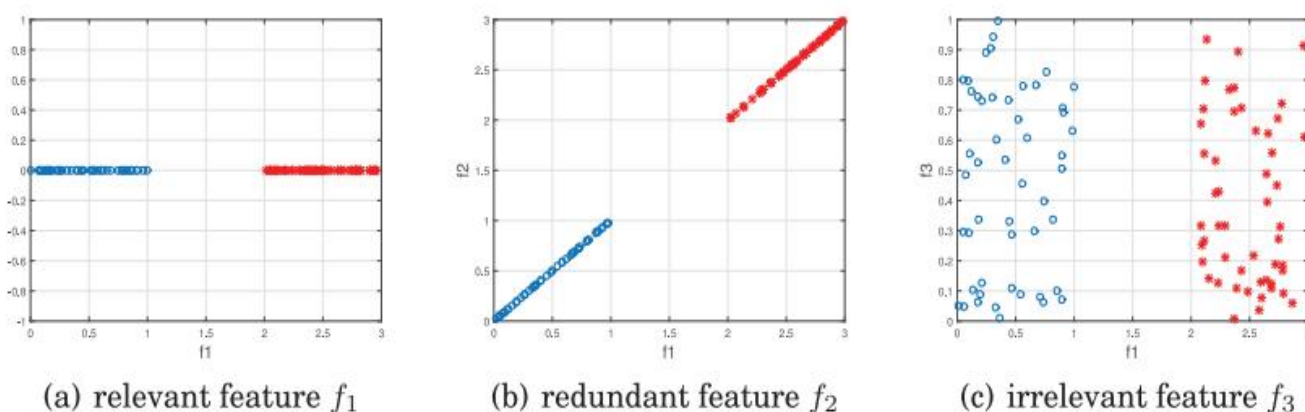


图 2-6:没有用的特征（引自[52]）

我们在经过特征抽取以后，一个动作序列变成了一个 13800 维的特征向量，毫无疑问这是一个非常高的特征数量，在论文当中我们使用了 2 种特征选择的方法：RFE 和基于方差的 Filter 方法。

(1) 基于 Wrapper 的方法：Wrapper 方法的前提是一个给定的机器学习算法来对所选的特征的有效性进行评分。一个属于 Wrapper 的特征提取方法一般都是要存在两个步骤的：

①通过某种方法来得到一个特征子集 ②对这个特征子集进行效用的评估

在这些基于 Wrapper 的特征选择方法当中要重复操作①和②，直到达到某种标准（例如特征已经达到最佳或者算法的迭代次数上限已经达到）才能停下来。然后把这一次搜索到的所有的特征拼接成一个特征子集，然后将学习算法作为一个黑盒，把特征子集对应的原数据的数据子集作为输入进去，然后根据算法的输出结果以及效果来对选取的特征子集的合理性进行判断。然后返回在预定的机器学习算法当中表现地最好的特征子集作为所选的特征。这种方法的思想虽然非常好，但是在实际应用的过程当中会遇到一个非常困难的问题就是：Wrapper 的时间复杂度实在是太高了。我们假设一个数据集的有着 n 个特征，那么如果要把对应的特征子集全部搜索完的话就要搜索 2^n 次，当 n 非常大时要把整个搜索空间都搜索完的话是非常不合实际的。所以如何制定搜索策略就是一个非常关键的问题。

我们采用了一个非常不错的搜索策略是递归特征消除法（RFE）。这个方法所使用的是一种贪心策略的方法。具体的步骤如下：

①先把原始的特征集 D 作为输入输进去机器学习算法，然后根据机器学习算法的效果对原有的特征根据某些属性（例如 `feature_importances_` 属性进行排序）

②剔除掉一些重要性比较低的特征，保留一些重要特征构成新的特征集 D^1 。

③ D^1 作为输入再次输入机器学习算法当中，进行和①②一样的操作，而后往复循环地进重复①②的操作，直到所选的特征数量达到要求缩减的数量为止。

在论文当中我们选择了随机森林作为 RFE 的给定的机器学习方法。

(2) 基于 Filter 的方法：这种方法依靠数据的一些具有统计学意义的量来评估特征的重要性。基于 Filter 分方法一般包含两个步骤：第一步是根据这些具有统计学意义的特征评价标准对特征的重要性进行排序。第二步就是低排名特征被过滤掉。

在论文当中我们采用了方差作为选定的具有统计学意义的量。方差反应的就是特征数据的变化情况，如果一个特征的方差特别小，那意味着这个特征的数据变化情况特别小。则在分类问题之中这个特征不会因为类别的变化而发生变化，也就是说这个特征不会为分类问题提供有价值的信息。

3 机器学习建模

机器学习是一门极其复杂的学科，它的内容涵盖了各种各样的学科知识例如概率论、统计学，运筹学，最优化理论，信息论等等。它是专门用来让机器学习人类的知识，模拟人类的思维，让机器产生人类的思维。这个学科是当今大火的人工智能的核心技术。自从计算机技术的蓬勃发展以来，机器学习模型也在发展。如果说前面的特征工程的好坏决定了数据挖掘任务的上限的话，那么建立越好的机器学习模型就会使得数据挖掘任务的效果就越逼近这个上限。在论文中，我们所要解决的问题就是根据上面抽取出来的特征对人体动作进行分类。所以对此我们建立了一个基于集成学习方法的分类模型，这个模型充分利用各种分类模型以及集成学习的优点，取得了不错的分类效果。

3.1 所用到的分类模型

分类模型是机器学习里面的一类非常经典的模型。分类模型的数据集都是带有对应的标签的，而机器学习分类模型的任务就是把数据正确归类到它所对应的那个标签去。分类模型从 70 年代开始到现在已经经过了几十年的发展，期间诞生出了许许多多的模型。在我们的论文当中我们使用了 6 种及其经典的分类模型，他们分别是：logistic 回归，支持向量机（SVM），KNN，随机森林，Xgboost，lightgbm。

3.1.1 logistic 回归

我们知道线性回归是一种非常经典的回归任务的机器学习方法。线性回归就是根据已知数据集里面的特征，训练出一个线性的回归方程。但是线性回归的输出结果是一个连续的结果，而分类问题需要的是一个可以输出离散的分类结果的分类器。所以线性回归不可以直接运用于分类作用。但是分类问题可以理解为是选择概率最大的类别作为分类器，而概率是一个 0 到 1 之间的连续值。所以通过某种映射使得线性回归的结果出现在 0-1 之间，那么就有可能把线性回归运用到分类问题中来，下图的 sigmoid 函数就是一个不错的选择：

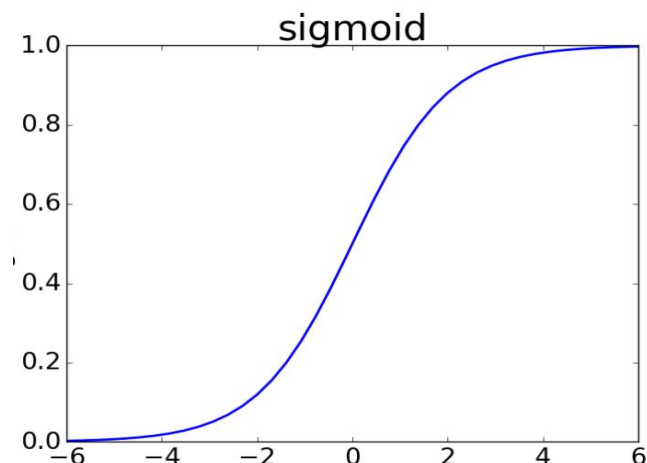


图 3-1: sigmoid 函数的图像

以二分类为例，分类器把某个样本点归类为“1”的根据就是分类器预测该样本点是“1”的概率至少大于 0.5。显然 sigmoid 函数就适合二分类的条件，这就使得这个函数十分适合用于分类问题的映射。所以研究者们就设计出了 logistic 二元分类器，具体的分类器函数如下公式所示：

$$\begin{aligned} P(Y=1|X=x) &= \frac{1}{1+e^{-(wx+b)}} \\ P(Y=0|X=x) &= 1-P(Y=1|X=x) \end{aligned} \quad (2)$$

为了找到最合适的 W 和 b ，我们采用的是极大似然估计法。极大似然估计法的核心思想就是找到适合的参数，让参与极大似然估计的样本出现的可能性最大。将其应用在二元分类的应用情况下就是：找到特定的 W 和 b ，使得所有样本都被 logistic 二元分类器正确分类的概率最大。所以研究者们设计 logistic 二元分类器的损失函数如下：

$$\prod_{i=1}^N [\pi(x)]^{y_i} [1-\pi(x)]^{1-y_i}, \quad \pi(x) = \frac{1}{1+e^{-(wx+b)}} \quad (3)$$

接着只要估计出使得似然函数最大的 w 和 b 就行了。logistic 多分类就是把的 logistic 二分类损失函数扩展到多分类的情况下就可以了。

3.1.2 支持向量机

支持向量机（SVM）属于监督学习，并且是机器学习里面最为经典并且数学推导最为优美的分类模型之一。支持向量机的核心理念就是训练出一个超平面，这个超平面是两个类别之间的最大间隔平面。这就把一个机器学习的参数估计问题转化成了一个最优化理论的凸优化问题，相比起 logistic 回归，支持向量机的推导过程中表现出了更加严谨的数学性以及更加良好的可解释性从数据的分布的角度来看，一个二分类问题的本质就是寻找一个可以把训练集中的属于不同类别的数据分割开来的超平面。但是有许多的超平面都可以做到这一点（如下所示）。所以可以肯定在这些超平面当中一定存在着某个超平面使得两边的分类的效果达到最好。SVM 就是旨在找到这个向量

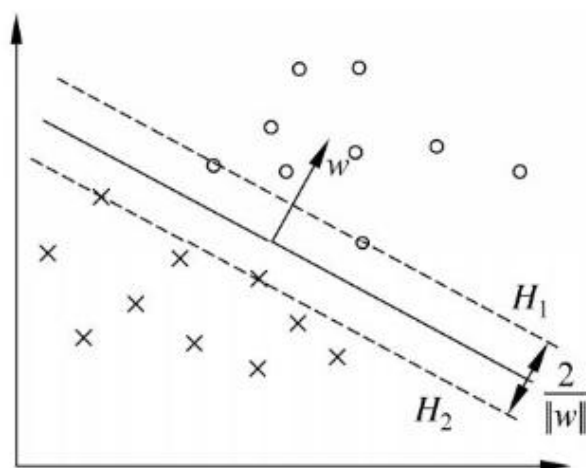


图 3-2：支持向量机效果图（引自[50]）

一般来说我们可以使用数据点到超平面的距离来对样本点属于某一个类的可能性进行量化统计。例如我们假设数据点 x_1 和 x_2 的数据点都是同一个类别的。如果 x_1 到超平面之间的距离小于 x_2 到超平面之间的距离，很明显 x_2 属于该类别的可能性会更加大。所以我们要找到的那一个“效果最好”的超平面就是满足让所有的数据点到超平面的距离之和达到最大。在这里研究者们引入了下面的几何间隔公式来进行量化：

$$r_i = \left| \frac{w}{\|w\|} \cdot x_i + \frac{b}{\|w\|} \right| \quad (4)$$

SVM 的核心就是在保持成功分类的前提之下使得所有样本点的几何间隔之和达到最大。这很明显是一个运筹学中的凸优化问题，最后可以运用拉格朗日对偶性采用对偶算法直接把 w 和 b 求出来。

上述是 SVM 进行二分类模型的效果。其实 SVM 也可以拓展到多分类模型的应用场合之下来的，具体的步骤可以分为 3 步：

- ①从训练数据集当中挑出某一类 A，将剩下的数据统一归为一类 B
- ②利用二元 SVM 的训练方法训练吃 A 和 B 之间的最大间隔超平面。
- ③重复步骤①②，直到训练出所有用于分类的超平面为止。

3.1.3 KNN

KNN 算法的思想是非常简单的：就是给定一个训练数据集，对于一个将要被分类的数据样本点，在数据集当中用某种方式找出数据集里面与这个实例“最近的” k 个样本点。然后通过统计看出哪一个类别出现的最多，就把这个样本点归为哪一类。从 KNN 算法的步骤可看出特征工程对 KNN 的表现起着很大的作用。

十分明显在 KNN 算法中 k 值的选取十分地重要。如果选择 k 值太小，就意味着该样本点的类别仅仅取决于周边的一个很小的领域的类别分布情况，这使得 KNN 模型会变得比较复杂。从而比较容易发生模型的过拟合。如果选择比较大的 k 值，那么模型就会兼顾

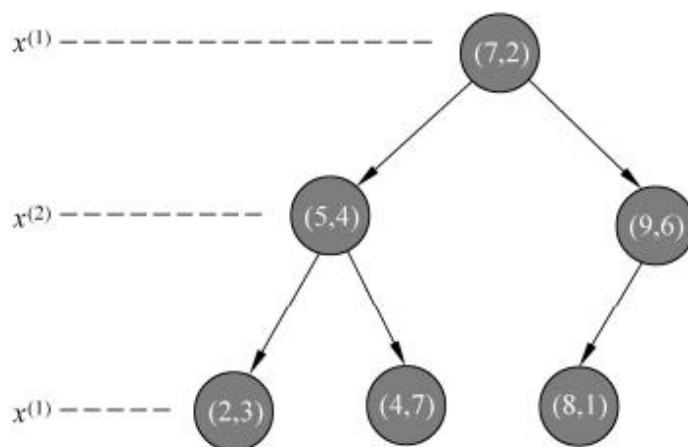


图 3-3: kd 树实例（引自[48]）

到那些距离样本点比较远的数据。可以比较客观地对样本点进行分类。但是这样就会导致 KNN 模型会比较简单，从而可能没有办法很好地对数据进行充分的拟合。

怎样在数据集当中搜索出 k 个样本点进行比对也是 KNN 算法的一个难题。最简单的实现方法是暴力线性扫描。但是这个方法的时间复杂度达到了 $O(n)$ 。所以研究学者们通常使用 kd 树（以图 3-3 为例子）这一特殊的存储结构来存储用于训练的数据。

3.1.4 基于树的分类算法

我们知道机器学习里面有一类算法是极其出名的：以树为基础的算法。这样的决策树一般都是由有向边和节点构成。节点有两种：非叶子节点代表数据集里面的某一个内部特征，叶子节点代表着数据集里面的某一个类别。

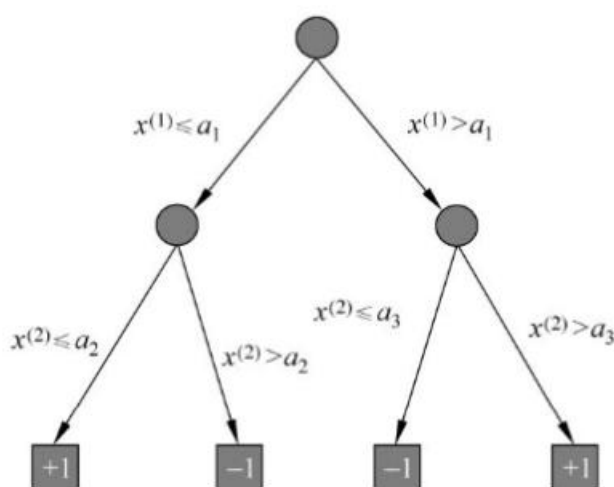


图 3-4：分类决策树实例（引自[49]）

将样本输入以后，先按照根节点里面的特征规则划定路径，然后一步步从根据非叶子节点的规则往下走，最后走到哪一个叶子结点就把该数据样本点划为对应的那个类别。所以在训练决策树的过程当中，选取特征就成为了十分重要的一步。ID3 算法的核心是利用特征的信息增益来选择特征，C4.5 算法则使用的是信息增益比，而 CART 算法使用的是基尼系数。

在之后的机器学习的发展中，研究者们提出了“集成学习”这一种机器学习策略。所以很多关于“树”的集成学习模型也被提了出来，下面简要介绍论文当中会使用到了 3 种模型：随机森林，xgboost 和 Lightgbm

（1）随机森林：这是决策树在 Bagging 的基础上诞生的模型。具体训练的过程如下：

- ①从数据集中随机选取 m 个数据样本点，从数据集特征之中随机选取 n 个特征
- ②利用①选取的数据与特征来对决策树进行训练。
- ③多次重复步骤①和②训练出多颗决策树

在实际的分类预测当中每一刻决策树都要参与分类，最后所有的分类预测结果中数量最多的那一个类别就是模型的分类结果。

(2) 梯度提升树: GDBT 是在决策树中的 CART 回归树的基础上使用 Boosting 集成策略所组成的模型。它的训练策略是: 训练多个学习器, 其中训练第 K 个基学习器模型的目的是为了拟合前 $K-1$ 个基学习器与真实值的残差, 最后模型的输出是所有学习器之和。而且在训练的过程当中使用了梯度下降法作为优化手段。GDBT 也可以用作分类问题。

Xgboost^[45]和 Lightgbm^[46]是 GDBT 的两个在工程上的非常优秀的实现。Xgboost 在优化过程当中使用了泰勒公式的二阶展开, 大量提升了算法的速度, 而且在优化的目标函数当中加入了正则化降低模型过拟合的风险。而 Lightgbm 则使用了单边梯度采样和互斥特征捆绑来解决数据集过大以及数据集特征过多的问题。

3.2 集成学习

集成学习方法是现在机器学习研究领域里面的一项十分重要的方法。集成学习方法的核心就是通过某种人为设计的策略, 可以同时把多个机器学习模型给充分利用起来, 从而达到一个取长补短, 让模型的预测效果变得更加优秀。

我们知道机器学习模型的模型复杂度和模型学习的效果是有一定关系的, 模型的复杂度越高, 那么说明模型的拟合能力就越强, 缺点就是很容易发生过拟合, 这样就使得训练出来的模型可以很好地拟合训练集, 但是模型没有很好的适配性。反之如果模型越简单, 所训练出来的模型对于非训练集数据的适配能力就会显著增强, 但是由于模型拟合能力不够的话就很容易造成模型的拟合效果很差。这个关系可以在下图得到一个反映:

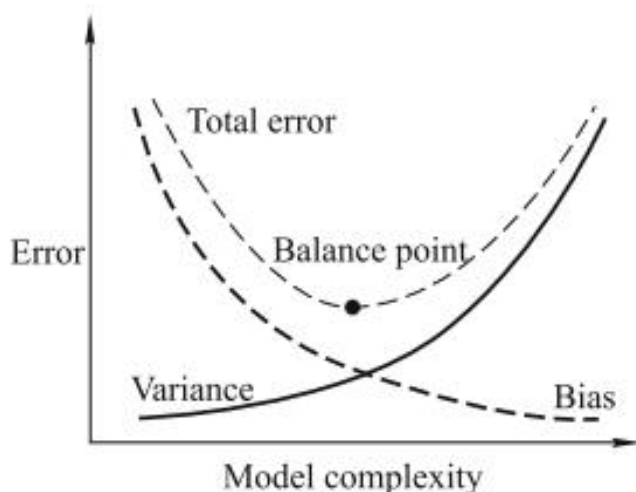


图 3-5 模型效果与模型复杂度的关系图（引自[47]）

可见当模型复杂度升高的时候。Bias 逐渐降低意味着模型的拟合能力逐渐提升, 而 Variance 却逐渐升高, 这使得模型的过拟合现象也逐渐地提升。所以一个模型的效果要综合的考虑到 Bias 和 Variance, 这就是 Total error。上述 Total error 图像证明了存在一个“平衡点”可以同时把 Bias 和 Variance 调配好, 这个点就是机器学习效果最好的点。所以能否成功进行集成学习的一个关键点就是: 利用怎样的集成学习策略把涉及到的机器学习模型组装成一个兼顾 Bias 和 Variance 的集成学习模型。集成学习的策略有许多种例如 bagging,

boosting 和 stacking 这几种。我们的论文当中仅仅涉及到了 bagging 和 boosting。

3.2.1 Bagging 算法

Bagging 集成学习首先是在上世纪 90 年代提出来的。Bagging 的思想是有放回地从训练集当中取出 M 个数据样本，依次放入多个机器学习模型当中进行训练，这样可以得到多个预测模型。对于分类问题，Bagging 所采取的策略是根据各个模型预测出分类结果以后投票表决产生最终的分类结果

Bagging 的核心思想综合考虑了所有模型的预测结果。因为不同的模型学习能力以及学习特点都是不一样的。综合考虑所有模型的预测结果可以说是起到了一个“取长补短”的效果，这样就使得模型对于训练数据集外数据的适配性大大增加，也使得过拟合现象发生的可能性以及过拟合的程度大大减小。可见 Bagging 方法是一种降低模型 Variance 的集成学习方法。

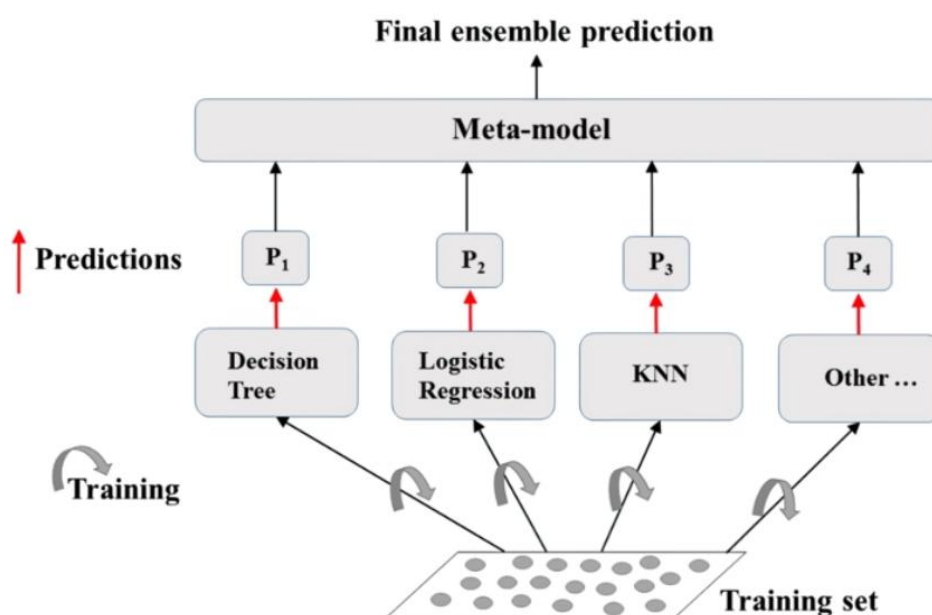


图 3-6: Bagging 算法的示意图（引自[51]）

3.2.2 Boosting 算法

Boosting 集成学习方法的思想是通过获得训练数据集之中的多个样本子集，训练出多个“弱机器学习模型”，而后基于这几个弱机器学习模型的综合预测结果产生出一个强机器学习模型。而且在这一轮轮训练的过程当中这些样本子集都是附带着权重的。

在分类问题当中。Adaboost 就是一个经典的 Boosting 算法的应用。在 Adaboost 算法的初始情况之中，所有的数据点在数据集当中的分布都是均匀的，也就是说一开始所有的数据样本点在模型训练当中都是起着相同的作用的。经过一轮训练以后总有一些错误分类的情况存在，那么此时为了将模型训练的注意力转移到那些分类错误的样本点上的时候，所有样本点的权重都会发生变化。分类正确的数据样本点在总的 Loss 函数上的权重会降低，而分类错误的数据样本点在 Loss 函数上面的权重则会提高。按照这个流程训练下去会得到多个分类器。如何综合利用那么多个分类器所得出来的结果又是一个问题。Adaboost 的策略就是每个分类器进行投票，票数多的那个类就是分类结果。

在回归问题当中，Boosting 的思想就是拟合模型输出和真实值的残差。先用原始的数据集训练出了第一个回归模型 M1，而后计算以下 M1 的输出和真实值的差距。而后再训练一个回归模型 M2 来拟合 M1 的输出和真实值的差距。然后这样不断地训练回归器。最后预测的时候就是把待预测的样本点输入多个回归模型当中进行预测。然后把所有的预测值加起来就是最终的预测结果。

我们可以看出来 Boosting 的算法是通过增加模型的复杂度来让很多的“弱机器学习模型”变成“强机器学习模型”，就是说这种方法让模型的拟合能力提高了。显然这是一种让模型的 bias error 变低的集成学习方法。

3.2.3 我们的机器学习模型

Boosting 让模型的 bias error 变低却让 Variance 变高，这就是说 Boosting 在增加了模型拟合能力的同时很容易让模型发生过拟合现象。同时 Bagging 又可以比较好地降低过拟合现象，基于这两种集成算法的优缺点，我们在论文中使用的集成策略综合了两种集成算法，具体的分类模型结构如下所示：

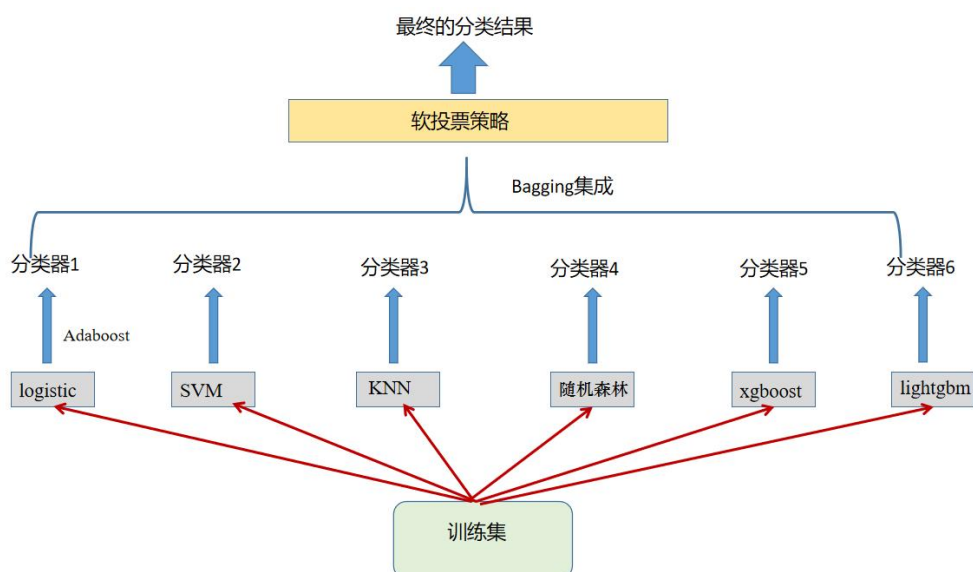


图 3-7：我们的集成学习模型

在我们的集成学习建模当中，我们先让 **logistic** 回归都根据已有的全部训练集为进行 **Adaboost** 来提高模型的拟合能力。因为这个模型的拟合能力相对其他模型比较差一些。**Lightgbm**，**Xgboost** 和随机森林都是已经经过集成学习策略优化的模型了所以不需要再次集成。而后使用 6 个分类器进行预测以后使用软投票策略得到最终分类结果。这样集成既考虑到了 **Boosting** 对模型拟合能力的提升也考虑到了 **Bagging** 对整个模型 **Variance** 的降低。找到了一个 **Bias** 和 **Variance** 的平衡点。

4 在 G3D 数据集上的实验

为了验证本文中提出的基于关节夹角的人体动作识别的方法，我们使用了 G3D 数据集作为实验的数据集。下面是对 G3D 的数据集的描述以及对本论文提出的方法的一系列的实验结果。

4.1. G3D 数据集

G3D^[44] 是 CMU 实验室收集的关于人体动作的数据集。该数据集包含 10 个受试者执行 20 个游戏动作：20 个游戏动作记录在 7 个动作序列中如下表所示所示。大多数序列包含多个动作，这些动作是在带有固定摄像头的受控室内环境中进行的。

表 1: G3D 数据集的包含关系

动作序列分类	游戏动作
第一人称射击	向中心射击,向右边射击,走,跑, 向左边射击 ,跳, 爬, 蹲
搏击	右拳, 左拳, 右踢, 左踢, 防守
高尔夫	挥杆
网球	正手, 反手, 发球
打保龄球	扔保龄球
驾驶	打方向盘,右转, 左转
混合动作	挥手, 拍打手臂,鼓掌

每一个受试者在执行动作的时候都要把动作序列重复执行 3 遍。并且在这个数据集当中每一个执行的动作都具有编号。数据集当中一共有 209 个动作序列。一共分为 7 类，具体的分布如下图所示：

Actor	Fighting				Golf			Tennis			Bowling			FPS		Driving			Misc		
1	22	23	24	25	26	27	28	29	30	31	32	33*	34	35	36	37	38	39	40	41	42
2	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
3	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84
4	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105
5	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126
6	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147
7	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168
8	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189
9	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210
10	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234

图 4-1: 动作序列编号归属

每一帧人体动作都包含 4 种数据：深度图，RGB 图片，转化为颜色坐标的深度图和骨架数据。我们的算法是基于 RGB 数据集的，所以论文实验只是用到了深度图和彩色图 2 种数据格式（如下图所示）。其中彩色图和深度图的存储格式为 640x480 的 PNG 文件。原始深度信息包含每个像素的深度。

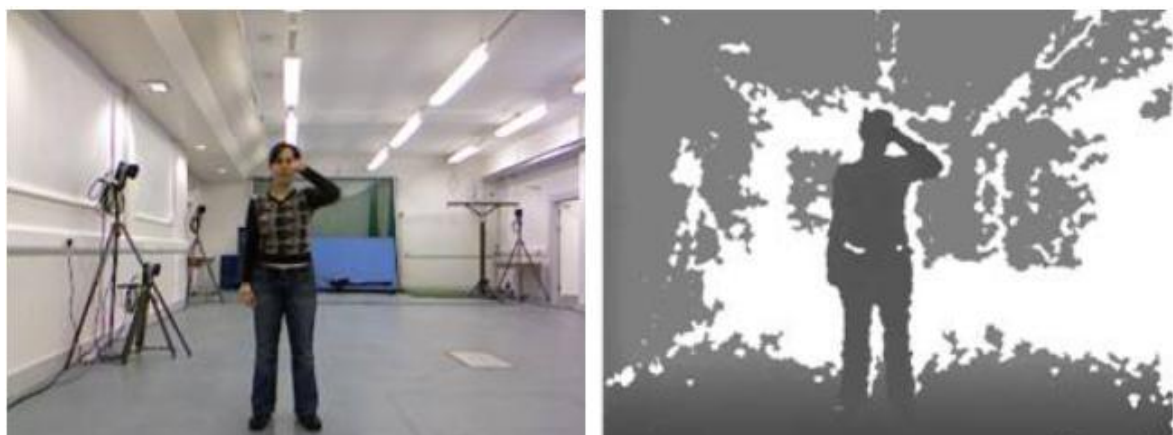


图 4-2：深度图和 RGB 图片

4. 2. Openpose 获取关键点的 2d 坐标

我们使用了 Openpose 的 Body25 预训练模型来对每一张图片进行关节点 2D 的坐标的提取。我们先是在 GitHub 上下载了 Openpose 的预训练模型。然后后使用了命令行窗口（cmd）来运行 Body25 预训练模型来获取关节点（如下所示）：

```
C:\Users\ASUS\Desktop\Openpose_2dtest>bin\OpenPoseDemo.exe --image
_dir examples\media\ --write_json .\output\ --display 0 --render
_pose 0
Starting OpenPose demo...
Configuring OpenPose...
Starting thread(s)...
Auto-detecting all available GPUs... Detected 1 GPU(s), using 1 of
them starting at GPU 0.
OpenPose demo successfully finished. Total time: 9.884127 seconds.
```

图 4-3：命令行窗口运行 Openpose

在 cmd 命令当中我们设置将输出的结果以 json 文件的形式输出在 output 文件夹里面，关节点输出在 json 文件的输出格式如下图所示：

```
{
  "version": 1.3,
  "people": [
    {
      "person_id": -1,
      "pose_keypoints_2d": [
        172.742, 50.4639, 0.965422, 172.734, 103.267, 0.896116, 135.226, 101.876, 0.785466, 118.593, 160.239, 0.874238, 185.279, 153.338, 0.787983, 214.371, 103.293, 0.808379, 208.887, 164.449, 0.746576, 146.341, 164.42, 0.691073, 168.586, 225.554, 0.614503, 139.42, 224.142, 0.616996, 128.253, 336.637, 0.74466, 126.894, 440.849, 0.774623, 194.956, 226.926, 0.660833, 200.518, 336.65, 0.646, 197.749, 439.46, 0.742315, 163.016, 40.7809, 0.914307, 182.507, 39.4533, 0.950236, 150.602, 47.6994, 0.875591, 195.005, 46.3176, 0.915796, 196.358, 482.552, 0.714264, 206.164, 479.744, 0.668328, 193.563, 445.019, 0.607059, 131.079, 479.765, 0.667279, 119.957, 474.218, 0.624634, 128.323, 447.791, 0.547799],
      "face_keypoints_2d": [],
      "hand_left_keypoints_2d": [],
      "hand_right_keypoints_2d": [],
      "pose_keypoints_3d": [],
      "face_keypoints_3d": [],
      "hand_left_keypoints_3d": [],
      "hand_right_keypoints_3d": []
    }
  ]
}
```

图 4-4: json 文件输出格式

我们可以看出 Openpose 的 2d 坐标的输出统一存储在 pose_keypoints_2d 的存储单元里面了。其中“0”号关节点的 2d 坐标为（172.742,50.4639），置信度为 0.965422。“1”号关节点的坐标为（172.734,103.267），置信度为 0.896116，依次到“24”号关节点。

4.3. 3D 坐标的引入以及动作特征向量的提取

众所周知，3D 坐标包含的信息比 2D 坐标多，所以我们认为在算法中使用 3D 关键点坐标取得的效果会比使用 2D 关键点坐标取得的效果更好。但是如果要通过 Openpose 获取 3D 坐标，就必须提供多个视角的图片而我们只有单个视角，所以我们只能通过 G3D 自带的深度数据集和关键点 2D 坐标来深度数据结合得到关键点的 3D 坐标。

为了验证这个想法，我们选取编号为“214”和“215”的两个动作序列，这是 Actor10 做的两遍“搏击”动作序列。然后，我们分别通过 2D 坐标和 3D 坐标抽取特征向量，而后计算余弦相似度作为两个动作序列之间的相似度。结果如下所示。基于 3D 坐标算法的相似度结果为 0.95147。但是计算基于二维坐标算法的相似度结果为 0.9484。显然，使用关键点的 3D 坐标来进行特征抽取是比较合理的。因此我们使用 3D 坐标来进行动作特征向量抽取，抽取出特征向量以后将其以 CSV 文件的格式存储起来。

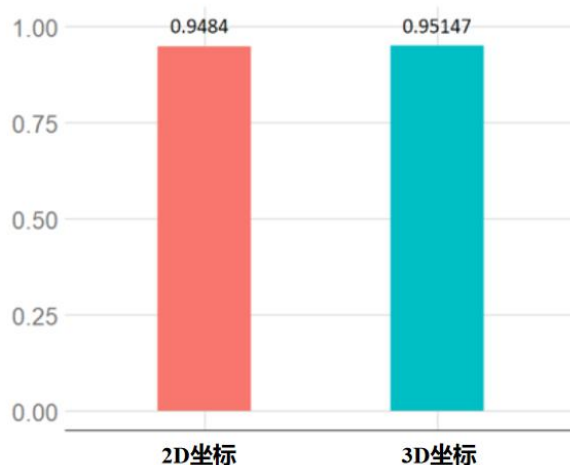


图 4-5: 2D 坐标和 3D 坐标的对比图

4.4. 特征选择

方差反应的就是特征数据的变化情况，如果一个特征的方差特别小，那意味着这个特征的数据变化情况特别小。则在分类问题之中这个特征不会因为类别的变化而发生变化，也就是说这个特征不会为分类问题提供有价值的信息。我们先用 `pandas` 库把所有经过抽取的特征向量集中到一起。而后我们通过 `pandas` 库来计算 13800 个特征的方差，计算结果如下图所示：

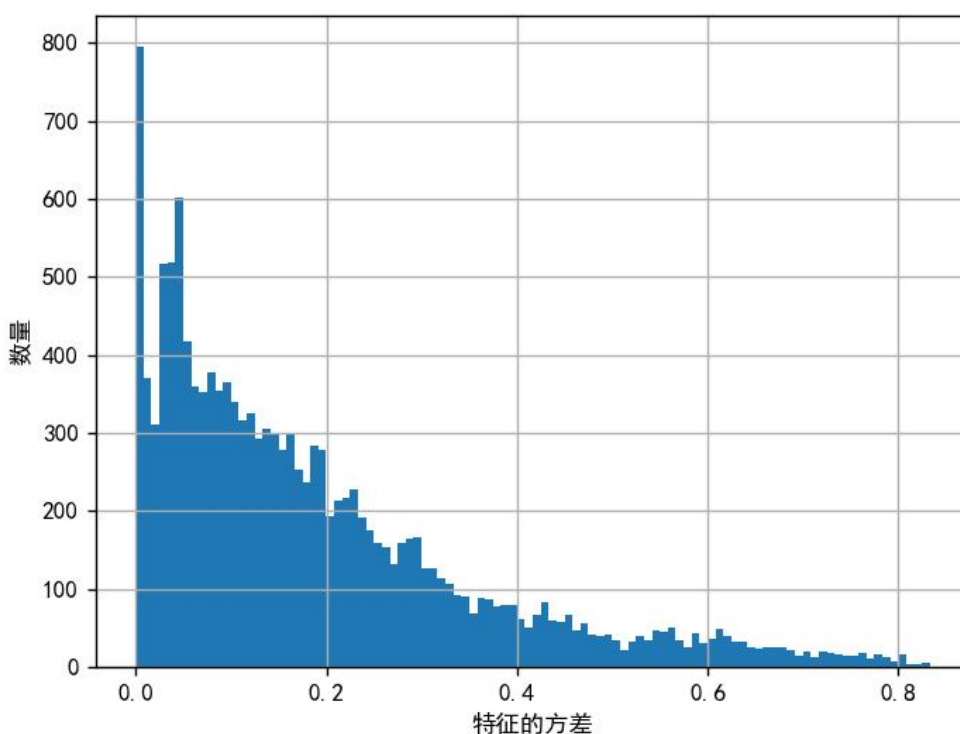


图 4-6： 抽取的特征的方差值

由上图可见方差在 0 到 0.2 的特征非常多。可见许多的特征没有什么变化，它们不会给分类任务提供非常多的有用的信息。所以我们有必要使用基于方差的 `Filter` 方法来先对特征进行第一轮选择。根据研究者们的做法方差低于 0.16 的特征都应该被剔除掉。所以我们将 0.16 作为参数进行 `Filter` 特征选择。

而后我们就使用 `sklearn` 的 `RFE` 模块来进行基于 `RFE` 的 `wrapper` 方法的特征选择。我们选择了随机森林来作为 `RFE` 的机器学习算法。最终数据集的特征的数量从 13800 降成了 500。

4.5. 机器学习分类模型调参训练

在本论文当中我们使用网格调参法作为参数选取的方法。机器学习库 `sklearn` 提供了一个专门用来进行网格调参法的类：`Gridsearch`，我们就使用这个库来进行调参。在实验当中我们使用的是 5 折模型验证。训练集与测试集的比值为 5:5。

4.5.1 logistic 回归

Logistic 回归的参数如下：

①penalty：这个参数就是用来防止过拟合。它的核心思想就是在 Loss 函数上面加上一个正则项，让损失函数在优化过程当中不要过度拟合。L1 代表选择绝对值范数作为正则项，L2 代表选择 2 范数作为正则项。在论文当中我们使用 L2。

C：正则化强度的倒数。C 越小对损失函数的惩罚越重，正则化的程度就越轻，对于过拟合现象的抑制效果就会更加明显，在“L2”的情况下随着 C 变化 Logistic 回归变化的情况如下图所示：

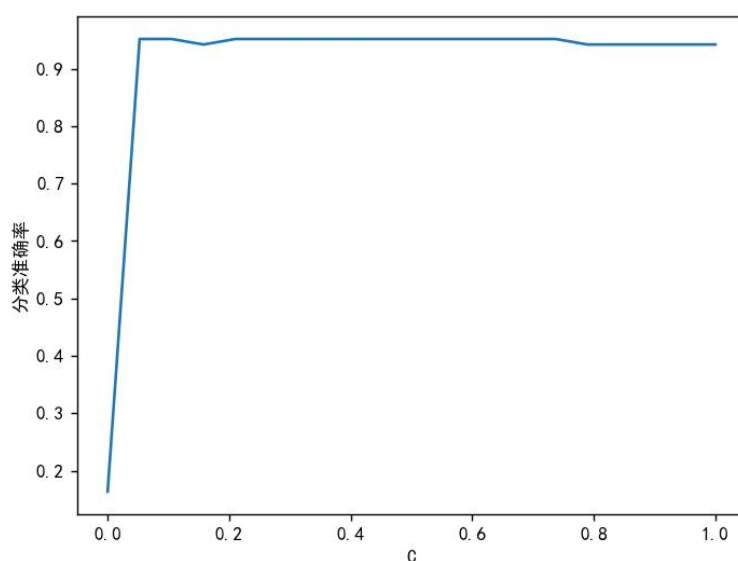


图 4-7： logistic 回归中 C 与识别率的关系

我们可以发现在 0 到 0.2 之间在上升，在 0.22 后没有怎么波动。在 0.65 以后往下跌了。通过网格化搜索的最佳参数为 0.214。

4.5.2 支持向量机

SVM 的参数如下：

①C：正则化强度的倒数。C 越小对损失函数的惩罚越重，正则化的程度就越轻，对于过拟合现象的抑制效果就会更加明显，从而对 SVM 中的“最大化函数间隔”的要求就会变得更加不那么严格。

②kernel(核函数)：在分类问题当中我们经常会遇到这样一种情况：在类别的交接区域当中会有不同类别的数据样本点混合在一起的情况。在这种情况下是不能找到一个超平面把类别之间分开来的。在这种情况下如果还想把 SVM 算法应用起来的话，就必须使用一个核函数把当前数据集映射到一个更加高维的空间里面，然后再那个高维空间里面执行 SVM 算法，而后把 SVM 求出来的超平面映射回来数据集原本的空间里面。Sklearn 中提供的常见的核函数有：poly，rbf（高斯核函数），具体的结果如下图所示：

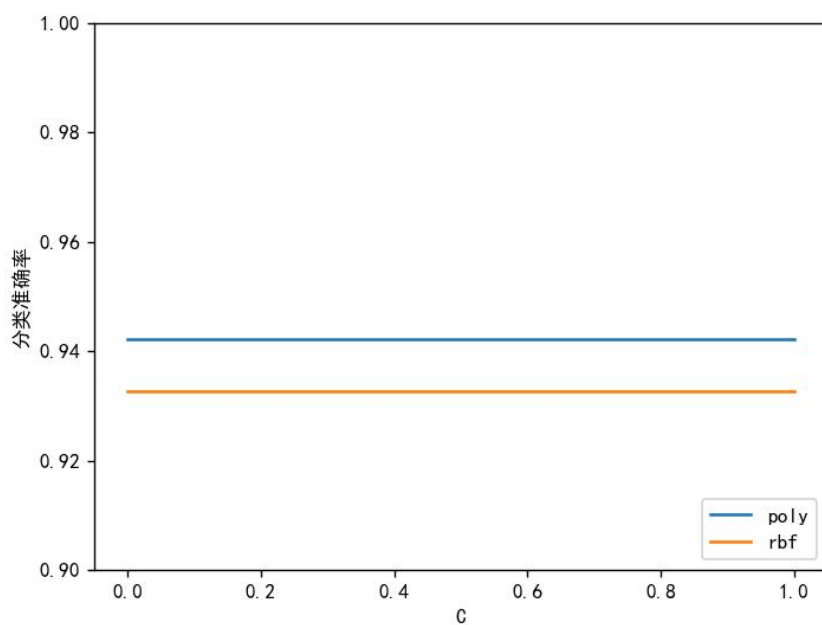


图 4-8: SVM 中 C 与分类模型识别率的关系

我们可以看出来使用 `poly` 作为 SVM 的核函数比较合理。并且 SVM 的准确率不随着学习率的变化而变化。所以在论文中我们设置 `kernel` 为 `poly`，网格搜索的最佳的 C 为 0.238。

4.5.3 KNN

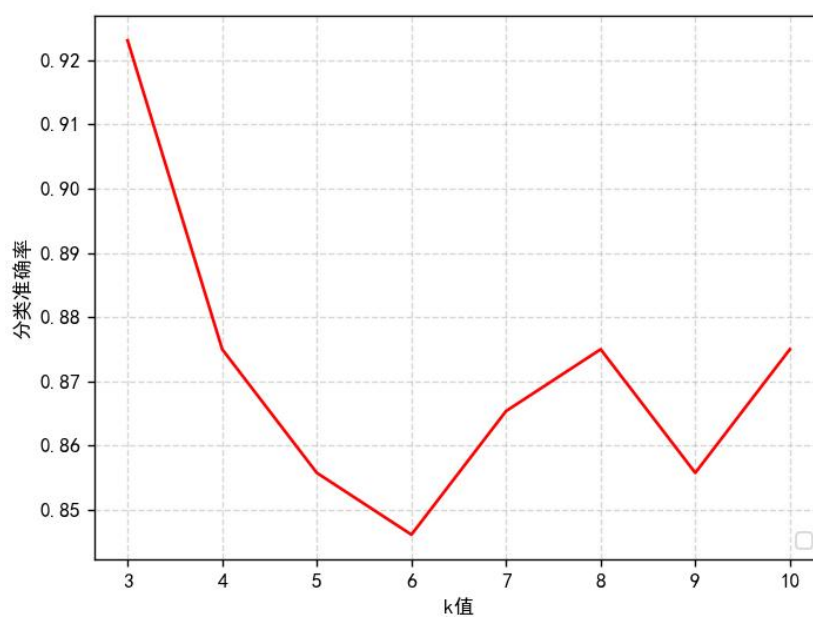


图 4-9: KNN 中 k 值与识别率的关系

KNN 算法里面有以下几种参数:

①n_neighbors: 就是 KNN 中的 k 值

②algorithm: 寻找 k 个样本点的方法。KNN 算法是选取 k 个距离待测样本点最近的 k 个样本点, 但如何寻找是很重要的, 如果选对了策略的话可以减少很多的时间。sklearn 所提供的算法提供了以下 kdtree 这种选项, 这种方法是非常省时的所以我们选择这种方法。

KNN 的分类结果的准确度随之 k 值的变化如上图所示。显然 K=3 的时候分类效果是最好的。此时 KNN 的分类准确率为 0.923。

4.5.4 随机森林

随机森林分类算法里面主要有以下参数:

①n_estimators: 随机森林算法里面包含的树的棵数。

②criterion: 这个参数就是随机森林里面的决策树在进行分裂的时候所遵从的标准, sklearn 为随机森林算法提供了以下几种标准:

(1) “gini”: 基尼系数 (2) entropy: 熵值

为了正确地选取参数我们做了实验, 效果如下所示:

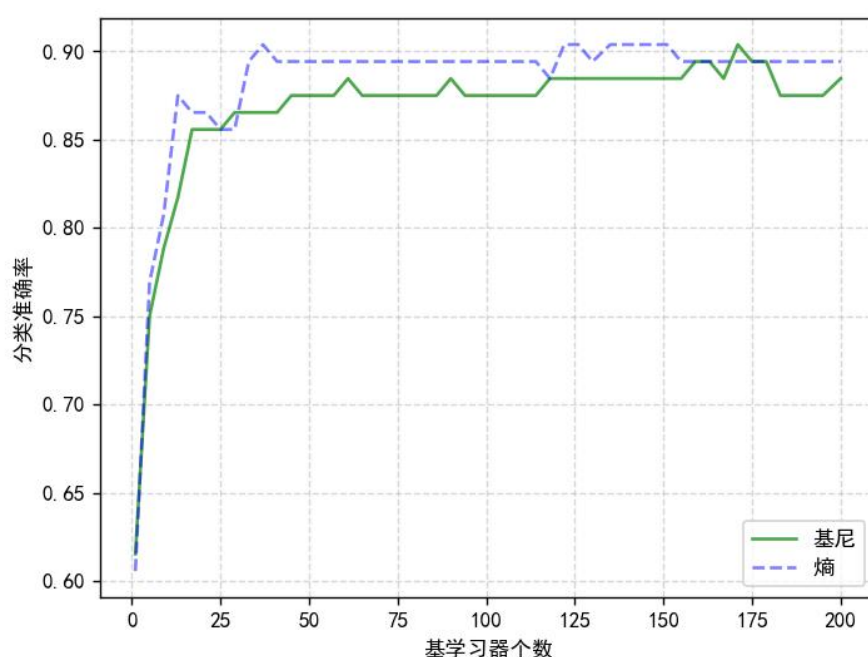


图 4-10: 随机森林中基学习器个数与识别率的关系

分析结果图可以知道 criterion 的两种参数当中平均最高的就是 “entropy” 这个选项。所以我们就选择 “entropy”。经过网格搜索以后最终我们得到 estimators 的最佳参数为 142, 此时的分类准确率为 0.9038

4.5.5 Xgboost 和 lightgbm

Xgboost 和 lightgbm 都是特别有用基于 Boosting 的集成模型，这两种模型是当今机器学习数据挖掘领域的相关比赛方面最火热的两个模型。所以我们也对这两种型进行了实验，结果如下：

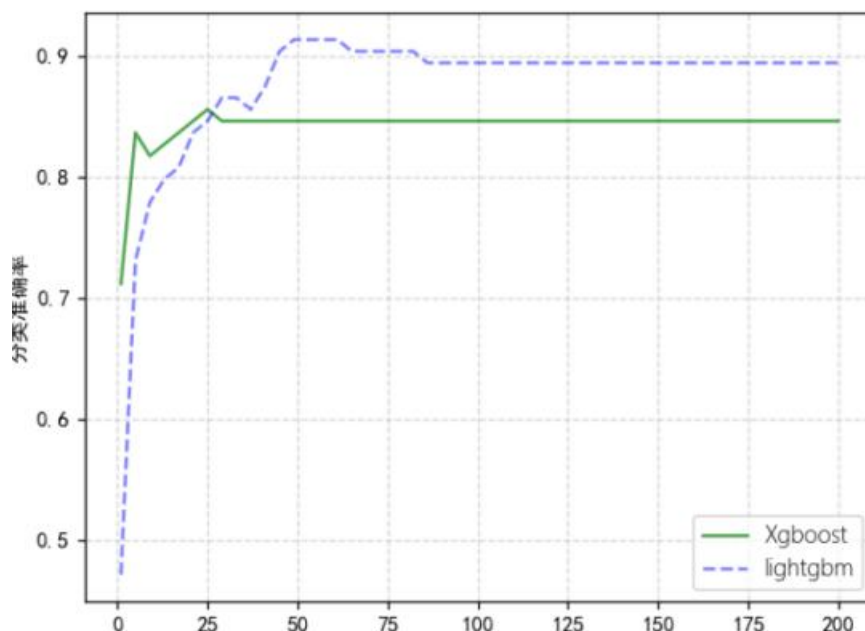


图 4-11: Xgboost 和 lightgbm 中 estimators 与识别率的关系

可以看出这两种模型都是随着 estimators 的增长，识别率达到峰值以后下降。经过网格搜索化以后得到的最佳参数为

Xgboost 的最佳 estimators 的个数为 25，此时识别率为 0.8557

Lightgbm 的最佳 estimators 的个数为 67，此时识别率为：0.9134

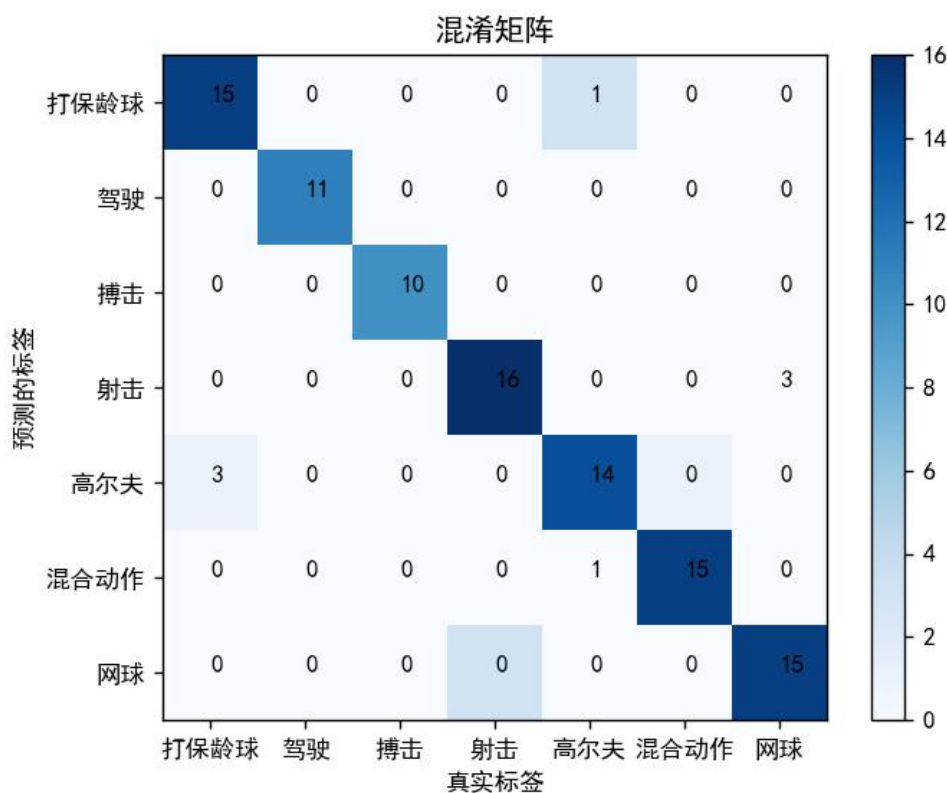
4.5.6 集成学习模型的分类效果

经过基本的调参以后，我们依次找到了我们的集成学习模型当中的每一个基分类器的最佳参数，搜索的结果如下所示：

- ①logistic 回归：“L2” 惩罚项，C 取 0.214
- ②支持向量机：核函数取 “poly”，C 取 0.238
- ③KNN：使用 kdtree 进行搜索，K 值取 3
- ④随机森林：estimators 的最佳参数为 142，criterion 选择 “entropy” 这个选项
- ⑤Xgboost：最佳 estimators 的个数为 25
- ⑥Lightgbm：最佳 estimators 的个数为 67

接下来我们就要对 logistic 回归进行 Adaboost 加强，在这里我们使用 Sklearn 库提供的 Adaboost 集成学习分类器。在 Adaboost 当中有一个十分关键的问题就是 Adaboost 中基分

类器的个数，因为这代表了集成学习方法的拟合能力大小。在实验当中基分类器的个数设定为 100。而后我们把所有的参数训练出来以后带入模型当中，而后把训练集测试集按照 5:5 的形式进行划分而后投入模型进行训练，最后得到的模型分类准确率达到了 92.3%，预测的分类结果的混淆矩阵如下图所示：



4-12 预测结果的混淆矩阵

从图中我们可以看出来“打保龄球”，“高尔夫”，“网球”各有 3 个数据样本点出现了预测错误。原因主要是因为 Openpose 在这三个动作里面的关节点识别准确度方面表现得不是十分优秀。

5. 结论

人体动作分类一直是一个非常复杂而且较为困难的问题。众多人体动作识别算法当中既有基于传统手工设计提取特征的传统机器学习方法，也有基于深度学习的端到端学习方法。但是两种方法都各有利弊，为此我们设计了一种结合了这两种方法的优点的算法：先用较为成熟的深度学习方法的预训练模型先从噪声比较多的人体动作数据之中把人体动作的关键点信息提取出来从而进一步提出人体关节夹角的信息，而后再使用传统的手工设计特征加上机器学习的方法进行建模。本文的主要的成果如下所示：

(1) 本文提出了一个基于人体关节夹角的人体动作序列特征提取方法。通过对以往的人体识别方法以及对人体分类算法的回顾，我们可以将其分为两种类型：基于人工设计的特征提取的分类方法以及深度学习方法。但是基于人工设计的特征提取的方法很容易受到一些视频数据以及图像数据本身所在来的噪声的影响。因为深度学习模型强大的拟合能力，所以基于深度学习的人体动作特征提取方法可以非常好的避免噪声的影响。但是与此同时深度学习方法的解释性要相对差一些，以及使用深度学习的方法所需要的算力以及对模型的存储所需要的存储空间都比较大。

为了克服这些问题，本论文提出了一个基于人体关节夹角的人体动作序列特征抽取方法：先用较为成熟的深度学习方法的预训练模型先从噪声比较多的人体动作数据之中把人体动作的关键点信息提取出来，从而进一步提出人体关节夹角的信息，而后再使用传统的手工设计特征加上机器学习的方法进行建模。这种方法既可以解决背景噪声比较多的问题，又具备良好的解释性，耗费的算力也不多。

在论文当中我们使用这个方法在 G3D 数据集上面以训练集:数据集为 5:5 的比例进行了测试。每一个基学习器的测试结果如下：

- ①Logistic+Adaboost 取得了 93.26%的分类准确率。
- ②SVM 多分类器使用 poly 核函数，C 为 0.238 取得了 93.26%的分类准确率。
- ③Xgboost 取得了 85.5%的分类准确率。
- ④KNN 取得了 92.3%的分类准确率。
- ⑤随机森林取得了 90.38%的分类准确率
- ⑥Lightgbm 取得了 91.34%的分类准确率。

从识别的结果可以看出我们所设计的特征提取方法可以比较好地对人体动作序列进行建模。究其原因主要是三点：第一点是这个方法先使用先使用对背景噪声影响比较小的深度学习模型先对数据进行了预处理，这样多可以使得论文中的算法可以比较少地收到背景噪声的影响。第二点就是这个特征选取方法是使用等间距采样以后，把每一个抽取的特征向量按照先后的顺序进行拼接，这种处理方法是充分地考虑到了人体动作序列的时间性。第三点就是处理单张图片的特征提取方法充分地考虑到了人体在完成动作的时候各个肢体之间的关系。我们知道人体在完成某个动作序列的时候各个肢体之间并不是独

立存在的，而文章之中所提出的计算所有肢体两两之间的夹角余弦值作为特征表示的方法就充分考虑到了这一点。

(2) 本文提出了一个把 Bagging 和 Boosting 有机的结合起来的集成学习策略。集成学习就是通过某种策略将多个模型集成起来，通过群体决策来提高决策准确率。集成学习首要的问题是选择什么样的学习器以及如何集成多个基学习器即集成策略。我们知道机器学习模型的模型复杂度和模型学习的效果是有一定关系的，模型的复杂度越高，那么说明模型的拟合能力就越强，缺点就是很容易发生过拟合，这样就使得训练出来的模型可以很好地拟合训练集，但是模型没有很好地适配性。反之如果模型越简单，所训练出来的模型对于非训练集数据的适配能力就会显著增强，但是由于模型拟合能力不够的话就很容易造成模型的拟合效果很差。因此我们在设计机器学习策略的时候就要充分考虑到模型的 bias 和 Variance 该如何找到一个不错的平衡点。

本论文里面所运用的集成学习策略同时使用了 Boosting 方法和 Bagging 方法，并且很好的把它们融合在了一起。我们知道机器学习模型的模型复杂度和模型学习的效果是有一定关系的，模型的复杂度越高，那么说明模型的拟合能力就越强，缺点就是很容易发生过拟合。在本文当中，对于拟合能力相对比较弱一点的 Logistic 分类模型，我们先使用 Adaboost 方法增强它的拟合能力之后，而后让它参与 Bagging 投票。这种策略同时兼顾了模型的 bias 和 Variance。最终本论文提出的方法在 G3D 上的表现为 92.3%，基本高于所有的基分类模型，稍低于表现最好的 SVM 多分类器，Logistic 回归。这样的测试结果也验证该集成学习策略的可行性。

对于本论文仍然有许多需要改进的地方，主要有以下两点：

(1) 这个集成学习策略的结果虽然高于一些分类模型，但是依然稍低于表现最好的 SVM 多分类器，Logistic 回归。所以我们依然需要去改进一下这个集成学习策略，让他可以达到更好地分类准确率。

(2) 从最后的混淆矩阵可以看出该方法在“打保龄球”，“高尔夫”，“网球”各有 2-3 个数据样本点出现了预测错误。原因主要是因为 Openpose 在这三个动作里面关键点识别准确度方面表现得不是十分优秀，甚至出现了关键点预测失败的情况。

(3) 从基学习器的分类准确率我们可以看出非树模型的 KNN，SVM，logistic 回归的学习效果要比属于树模型的随机森林，xgboost，Lightgbm 要好一些。可以看出这种基于关节夹角的特征提取方法比较适用于非树模型的学习，所以我们后面可以对这项特征学习方法进行改进使它也变得适合树模型的学习

致谢

时间过得飞快，转眼之间大学就要结束了，在汕头大学的四年的学习以及生活时光使我受益匪浅。经历几个月的工作以后，终于完成了，想想这几个月以来的工作，有不少的人都对这次毕业论文的完成帮助极大，此刻要向他们表达我最诚挚的谢意。

首先，我要深深感谢我的导师赵志丹老师。赵老师为人谦和，平易近人。在论文的选题、搜集资料和写作阶段，赵老师都不但给予了十分充分的指导，还给予了莫大的关怀以及鼓励，让我更有信心地完成这次毕业论文。在论文的写作过程中，每当我遇到了什么无法解决的困难的时候，赵老师总会放下自己手头上的工作来对我进行指导；在我初稿完成之后，赵老师又在自己忙碌的时候仍然抽出时间为我进行指导。在指导的过程中赵老师十分细致的看我的论文，对于每一句话每一个图表他都要充分地过一遍。这也让我发现了许多自己的问题，也对这次毕业论文的完成起到了很大的作用。在这里，我再次向赵老师表达我的敬意，祝赵老师以后工作顺利，家庭幸福，论文发到手软。

同时也感谢我的表弟，以极大的热情，帮忙我完成了第一手数据集的收集，并且在数据集的搜索过程当中他反复地与我进行沟通以及交流，中间还发生过几次争执，不过这一切都是为了让我的毕业论文可以更好地完成。

最后，我要感谢参与我论文评审和答辩的各位老师，他们给了我一个审视几年来学习成果的机会，让我能够明确今后的发展方向，他们对我的帮忙是一笔无价的财富。我将在今后的工作、学习中加倍努力，以期能够取得更多成果回报他们、回报社会。再次感谢他们，祝他们一生幸福、安康！

参考文献

- [1]董九庆.(2021).基于视频和骨架数据的行为识别算法研究(硕士学位论文,上海工程技术大学).<https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD202301&filename=1022615260.nh>.
- [2]王妍,詹雨薇,罗昕,等.(2023).视频片段检索研究综述. 软件学报 (02),985-1006. doi:10.13328/j.cnki.jos.006707.
- [3]Papadopoulos, Axenopoulos, Daras. Real-time skeleton-tracking-based human action recognition using kinect data. In Proceedings of the International Conference on Multimedia Modeling, Dublin, Ireland,6–10 January 2014; pp. 473–483.
- [4]Paul, S. N, Singh, etal. (2014). Survey on video analysis of human walking motion. International Journal of Signal Processing, Image Processing and Pattern Recognition, 7(3), 99-122.
- [5]Yang, Tian, (2014). Super normal vector for activity recognition using depth sequences. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 804-811).
- [6]Li, Leung, H, Shum. Human action recognition via skeletal and depth based feature fusion.In Proceedings of the Motion in Games 2016, Burlingame, CA, USA, 10–12 October 2016; pp. 123–132.
- [7]Guler, R. A., Neverova, N., etal. (2018, June). Dense human pose estimation in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA (pp. 18-23).
- [8]Fang, H. S., Xie, S., etal(2017). Rmpe: Regional multi-person pose estimation. In Proceedings of the IEEE international conference on computer vision (pp. 2334-2343).
- [9]Cao, Z., Simon, T., Wei, etal (2017). Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7291-7299).
- [10]Papadopoulos, G.T, Axenopoulos, etal. Real-time skeleton-tracking-based human action recognition using kinect data. In Proceedings of the International Conference on Multimedia Modeling, Dublin, Ireland,6–10 January 2014; pp. 473–483.
- [11]Li, M,Leung,etal. Human action recognition via skeletal and depth based feature fusion.In Proceedings of the Motion in Games 2016, Burlingame, CA, USA, 10–12 October 2016; pp. 123–132
- [12]韩磊, 李君峰, 贾云得. (2010). 基于时空单词的两人交互行为识别方法. 计算机学报, 33(4), 776-784.
- [13]Zhang, H. B., Lei, etal (2016). Multi-surface analysis for human action recognition in video. SpringerPlus, 5, 1-14.

- [14]付朝霞, 王黎明. (2013). 基于时空兴趣点的人体行为识别. 微电子学与计算机, 30(8), 28-30.
- [15]王博. (2014). 基于时空兴趣点的人体行为识别方法研究 (Master's thesis, 南京邮电大学).
- [16]Burghouts, G. J., Schutte, K.,etal. (2014). Instantaneous threat detection based on a semantic representation of activities, zones and trajectories. *Signal, Image and Video Processing*, 8, 191-200.
- [17]Wang, H., Schmid, C. Action recognition with improved trajectories. In *Proceedings of the ICCV, Sydney,NSW, Australia*, 1–8 December 2013; pp. 3551–3558.
- [18]C. Schödl, I. Laptev, B. Caputo, “Recognizing human actions: Alocal svm approach,” in *IEEE ICPR*, 2004.
- [19]M. Blank, L. Gorelick, E. Shechtman,etal, “Actionsas space-time shapes,” in *Proc. ICCV*, 2005.
- [20]P. Dollar, V. Rabaud, G. Cottrell, etal “Behavior recogni tion via sparse spatio-temporal features,” in *VS-PETS*, 2005.
- [21]Yuan, H., Wang, C. (2011). A human action recognition algorithm based on semi-supervised kmeans clustering. *Transactions on edutainment VI*, 227-236.
- [22]Sekma M., Mejdoub M., Amar,etal. (2015). Human action recognition based on multi-layer fisher vector encoding method. *Pattern Recognition Letters*, 65, 37-43.
- [23]Q. Shi, L. Cheng, L. Wang, etal, “Human action segmenta tion and recognition using discriminative semi-markov models,” *IJCV*,vol. 93, pp. 22–32, 2011.
- [24]M. Ryoo , J. Aggarwal, “Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities,” in *ICCV*, 2009, pp. 1593–1600.
- [25]N. Ikizler , D. Forsyth, “Searching video for complex activities with finite state models,” in *CVPR*, 2007.
- [26]S. B. Wang, A. Quattoni, L.-P. Morency, etal,“Hidden conditional random fields for gesture recognition,” in *CVPR*,2006.
- [27]Z. Wang, J. Wang, J. Xiao, etal, “Substructural and boundary modeling for continuous action recognition,” in *CVPR*,2012.
- [28]Wu, X., Xu D.,etal. (2011, June). Action recognition using context and appearance distribution features. In *CVPR 2011* (pp. 489-496). IEEE.
- [29]Bui, N. N., Kim,etal (2015, July). Human action recognition based on GMM-UBM supervector using SVM with non-linear GMM KL and GUMI. In *Seventh International Conference on Digital Image Processing (ICDIP 2015)* (Vol. 9631, pp. 327-333). SPIE.
- [30]Yuan, C., Li,etal. (2013). Modeling geometric-temporal context with directional pyramid co-occurrence for action recognition. *IEEE Transactions on Image Processing*, 23(2), 658-672.

- [31]C. Fanti, L. Zelnik-Manor, P. Perona, “Hybrid models for human motion recognition,” in CVPR, 2005
- [32]J. C. Niebles, L. Fei-Fei, “A hierarchical model of shape and appearance for human action classification,” in CVPR, 2007.
- [33]S.-F. Wong, T.-K. Kim, and R. Cipolla, “Learning motion categories using both semantic and structural information,” in CVPR, 2007.
- [34]J. C. Niebles, H. Wang, L. Fei-Fei, “Unsupervised learning of human action categories using spatial-temporal words,” *International Journal of Computer Vision*, vol. 79, no. 3, pp. 299–318, 2008.
- [35]Y. Wang ,G. Mori, “Learning a discriminative hidden part model for human action recognition,” in NIPS, 2008.
- [36]Han, L., Wu, X., Liang, W., etal. (2010). Discriminative human action recognition in the learned hierarchical manifold space. *Image and Vision Computing*, 28(5), 836-849.
- [37]L. Wang , D. Suter, “Recognizing human activities from silhouettes:Motion subspace and factorial discriminative graphical model,” in CVPR, 2007.
- [38] K. Jia ,D.-Y. Yeung, “Human action recognition using local spatio temporal discriminant embedding,” in CVPR, 2008.
- [39]Y. Kong, Y. Jia, Y. Fu, “Interactive phrases: Semantic descriptions for human interaction recognition,” in PAMI, 2014.
- [40]Kong, Y., Jia, etal. (2012). Learning human interaction by interactive phrases. In *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision*, Florence, Italy, October 7-13, 2012, Proceedings, Part I 12 (pp. 300-313). Springer Berlin Heidelberg.
- [41]A. Vahdat, B. Gao, M. Ranjbar, etal, “A discriminative key pose sequence model for recognizing human interactions,” in *ICCV Workshops*, 2011, pp. 1729 –1736.
- [42]Wei, S. E., Ramakrishna, etal. (2016). Convolutional pose machines. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (pp. 4724-4732).
- [43]吴逸南,周炜业,罗泽龙,赵志丹.(2021).基于 OpenPose 的动作序列相似度计算算法..(eds.) 第八届中国体能高峰论坛暨第二届中国体能训练年会书面交流论文集(pp.865-876)..
- [44]V. Bloom, V. Argyriou , D. Makris, "Hierarchical transfer learning for online recognition of compound actions", *Computer Vision and Image Understanding*, vol. 144, pp. 62-72, 2016.
- [45]Chen, T., Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- [46]Ke, G., Meng, Q., Finley, etal. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
- [47]Dong, X., Yu, etal. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, 14, 241-258.

- [48]李航 . 统计学习方法[M]; 清华大学出版社;2012; p43
- [49]李航 . 统计学习方法[M]; 清华大学出版社;2012; p57
- [50]李航 . 统计学习方法[M]; 清华大学出版社;2012; p103
- [51]leon_kbl ;集成学习 Bagging ;2020.03.29 ; <https://www.jianshu.com/p/00ad8069e9ab>
- [52]Li J, Cheng K, Wang S, et al. Feature selection: A data perspective[J]. ACM computing surveys (CSUR), 2017, 50(6): 1-45.

汕头大学本科毕业论文（设计）评分表 (指导教师用)

论文（设计）题目		基于人体关节夹角的人体动作识别算法				
专业年级		19大数据	学生姓名	吴逸南	学号	2019612039
序号	评价项目	评价标准			项目得分 (填入整数分值, 最高分5, 最低分1)	
1	选题	与毕业论文（设计）内容相符；大小、深浅适度，份量合理；符合专业培养目标，体现综合训练的基本要求，具有一定的理论意义或实际价值。			5	
2	工作态度	态度认真，积极思考问题，主动采取措施解决问题，按期完成各项工作，及时提交有关材料。			5	
3	工作量	完成论文（设计）所付出的时间、精力、工作量符合有关要求或期望。			5	
4	写作水平	层次分明，结构合理；论述有据，条理清晰；表述流畅，用语准确；分析严谨，结论可靠；符合汕头大学《本科毕业论文撰写规范》各项要求。			4	
5	知识能力	系统地掌握了本专业相关的基础理论、专业知识、基本研究方法和技能；文献收集阅读广泛，综述能力强，了解本课题国内外的研究状况及最新的进展情况。			4	
6	学术水平与创意	论文有独到的见解，富有新意或对某些问题有较深刻的分析，有一定的学术水平或实用价值。			4	
平均分					4.5	
7	参考论文终稿检测结果说明论文是否涉嫌抄袭(空间不够时可另附页)：					
评语（请结合论文的具体内容进行个性化评述，指出论文的优点和不足之处）： 论文写作逻辑清晰，行文逻辑也比较合理，论文中提出的特征提取方法比较具有新意。但是论文写作格式存在一些问题。						
建议成绩评为： <u>优秀</u>						
指导教师（签字）： <u> </u> 日期：2023年5月5日						

注：1. 建议成绩的记分采用五级制，分为优秀、良好、中等、及格、不及格五个等级。建议成绩的评定请参考平均分及是否发现抄袭等情况得出，评分参考标准为：优秀：平均分 ≥ 4.5 ；良好： $3.5 \leq \text{平均分} < 4.5$ ；中等： $2.5 \leq \text{平均分} < 3.5$ ；及格： $1.5 \leq \text{平均分} < 2.5$ ；不及格：平均分 < 1.5 。发现抄袭情况时，可视抄袭程度轻重降低相应评价等级，情况严重者评为不及格。

2. 本表限一页。


汕头大学本科毕业论文（设计）评分表（评阅人用）

论文（设计）题目	基于人体关节夹角的人体动作识别算法				
专业年级	19 大数据	学生姓名	吴逸南	学号	2019612039
序号	评价项目	评价标准			项目得分 (填入整数分值, 最高分5, 最低分1)
1	选题	与毕业论文（设计）内容相符；大小、深浅适度，份量合理；符合专业培养目标，体现综合训练的基本要求，具有一定的理论意义或实际价值。			5
2	工作量	完成论文（设计）所付出的时间、精力、工作量符合有关要求或期望。			5
3	写作水平	层次分明，结构合理；论述有据，条理清晰；表述流畅，用语准确；分析严谨，结论可靠；符合汕头大学《本科毕业论文撰写规范》各项要求。			4
4	知识能力	系统地掌握了本专业相关的基础理论、专业知识、基本研究方法和技能；文献收集阅读广泛，综述能力强，了解本课题国内外的研究状况及最新的进展情况。			5
5	学术水平与创意	论文有独到的见解，富有新意或对某些问题有较深刻的分析，有一定的学术水平或实用价值。			4
平均分					4.5
6	参考论文终稿检测结果说明论文是否涉嫌抄袭(空间不够时可另附页)：				
评语：					
论文的逻辑比较合理，描述也比较清晰。目前尚存在以下问题，建议进行相应的修改。 (1) 论文的英文摘要存在部分的语法错误，建议修改。 (2) 目录的格式存在问题，请参考计算机系的要求进行调整。 (3) 图2-1比较模糊，建议调整为清晰版。 (4) 表2-3前面，没有表2-1和2-2，建议调整。 (5) 第3章，需要另起一页，建议调整。 (6) 第15、27、28、29页下方存在较大的空白，建议调整。 (7) 部分图过大，例如图4-7，建议调小。 (8) 部分参考文献的格式不规范，请参考计算机系的要求进行调整。					
建议成绩评为： <u>优秀</u>					
评阅人（签字）： <u>云萍</u> 日期：2023年5月6日					

注：1. 建议成绩的记分采用五级制，分为优秀、良好、中等、及格、不及格五个等级。建议成绩的评定请参考平均分及是否发现抄袭等情况得出，评分参考标准为：优秀：平均分 ≥ 4.5 ；良好： $3.5 \leq \text{平均分} < 4.5$ ；中等： $2.5 \leq \text{平均分} < 3.5$ ；及格： $1.5 \leq \text{平均分} < 2.5$ ；不及格：平均分 < 1.5 。发现抄袭情况时，可视抄袭程度轻重降低相应评价等级，情况严重者评为不及格。

2. 本表限一页。

汕头大学本科毕业论文（设计）评分表（答辩委员会用）

论文（设计）题目	基于人体关节夹角的人体动作识别方法		
专业年级	19数据科学与大数据技术	学生姓名	吴逸南
		学号	2019612039
序号	评价项目	评价参考标准	项目得分 (填入整数分值, 最高分5, 最低分1)
1	选题	与毕业论文（设计）内容相符；大小、深浅适度，份量合理；符合专业培养目标，体现综合训练的基本要求，具有一定的理论意义或实际价值。	5
2	知识能力	系统地掌握了本专业相关的基础理论、专业知识、基本研究方法和技能；文献收集阅读广泛，综述能力强，了解本课题国内外的研究状况及最新的进展情况。	5
3	学术水平与创意	论文有独到的见解，富有新意或对某些问题有较深刻的分析，有一定的学术水平或实用价值。	4
4	答辩表现	能简明、清楚地介绍论文的主要内容；能准确流利地回答他人所提问题。	5
平均分			4.8
5	参考论文终稿检测结果说明论文是否涉嫌抄袭(空间不够时可另附页)： 否		
答辩委员会意见： <input checked="" type="checkbox"/> 通过答辩，成绩达到学士学位论文水平。 <input type="checkbox"/> 未通过答辩，理由：			
最终成绩评定为：____优秀____			
答辩委员主任（签字）：  日期：2023年5月11日			

注：1. 最终成绩的记分采用五级制，分为优秀、良好、中等、及格、不及格五个等级，由答辩委员会参考本表平均分、是否发现抄袭情况并参考指导教师、评阅人意见得出。

2. 本表平均分与五级制的折合关系为：优秀：平均分 ≥ 4.5 ；良好： $3.5 \leq$ 平均分 < 4.5 ；中等： $2.5 \leq$ 平均分 < 3.5 ；及格： $1.5 \leq$ 平均分 < 2.5 ；不及格：平均分 < 1.5 。发现抄袭情况时，可视抄袭程度轻重降低相应评价等级，情况严重者评为不及格。

3. 本表限一页。