中国科学: 信息科学 2020 年 第 50 卷 第 3 期: 438–460
SCIENTIA SINICA Informationis
《中国科学》杂志社
SCIENCE CHINA PRESS
论文

一种基于时间序列特征的可解释步态识别方法

施沫寒, 王志海*

北京交通大学计算机与信息技术学院, 北京 100044

* 通信作者. E-mail: zhhwang@bjtu.edu.cn

收稿日期: 2018–12–21; 修回日期: 2019–04–01; 接受日期: 2019–04–28; 网络出版日期: 2020–03–03

国家自然科学基金 (批准号: 61672086, 61702030, 61771058)、北京市自然科学基金 (批准号: 4182052) 和中央高校基本科研业务费专项资金 (批准号: 2017YJS036) 资助项目

摘要 步态特征识别是生物特征识别的一种, 在大量实际场景中有广泛的应用. 目前, 基于深度学习的方法在步态识别任务中表现出较好的准确率. 但是, 在对机器学习的研究中, 人们不仅希望得到精确的预测, 还希望算法对识别结果进行解释, 以便人们理解实际问题中的关键. 深度神经网络的黑盒属性使得解释其识别依据非常困难. 在已有的步态识别文献中, 关注可解释性的研究尚处于空白状态. 另外, 深度神经网络需要大量数据来学习模型参数, 在问题规模较小时难以有效地在未见数据上泛化. 本文探索了一种兼具准确性和可解释性的步态识别方法. 将步态特征表示为多维时间序列, 使用一种基于 Shapelet 的时间序列分类方法进行步态识别. Shapelet 是时间序列中最具有辨别性的子序列, 基于 Shapelet 的时间序列分类方法能够提供较好的可解释性, 同时可以提供较高的准确率. 我们在 CASIA-B 数据集上进行了实验, 和几种较新的深度学习方法进行了比较. 实验表明, 本文提出的方法在较小规模的数据集上能够提供与深度神经网络接近的准确率. 与此同时, 还能详细具体地解释模型的决策依据, 即哪些特征在视频哪几帧的表现对某个个体而言最具辨别性.

关键词 步态识别, 时间序列, Shapelet, 随机森林, 可解释性

1 引言

生物特征识别是指使用内在的身体或行为特征辨别人的身份 [1]. 人体步态识别是其中的一种, 这种技术通过分析不同人的行走姿势来辨别人的身份. 尽管步态特征不像面部特征、指纹特征一样具有丰富的特征和充分的辨别性, 但由于其具备另外的优势, 使得步态识别技术成为具有广阔前景的下一代身份识别技术. 这些优势包括: 不需要会带来不便的实际身体接触; 容易获得用于识别的数据; 对图像分辨率等数据质量的要求低; 对于距离等客观条件的要求低等 [2]. 例如, 在智能家居应用中, 由于步态识别不需要任何动作配合, 可以全时段监控家庭人员的状态, 而不会对家庭生活带来任何不便. 在安防应用中, 不同的衣物和饰品会导致外观差异, 但这不会改变步态.

引用格式: 施沫寒, 王志海. 一种基于时间序列特征的可解释步态识别方法. 中国科学: 信息科学, 2020, 50: 438–460, doi: 10.1360/N112018-00326
Shi M H, Wang Z H. An interpretable gait recognition method based on time series features (in Chinese). Sci Sin Inform, 2020, 50: 438–460, doi: 10.1360/N112018-00326

© 2020《中国科学》杂志社

基于平面图像的步态识别技术大致可以分为两种: 基于模型的方法和基于轮廓的方法. 基于模型的方法通常会对每一帧的图像建模以估计人体结构信息 [2]. 尽管基于模型的方法可以较好地描述步态运动, 但是它的计算代价较高, 同时对图像分辨率的要求也较高. 基于轮廓的方法采用背景减除等方式抽取步态轮廓, 基于轮廓特征进行识别. 由于人的步态轮廓本身十分简单, 同时包含了大部分步态信息, 并且抽取该轮廓的过程对图像质量要求很低, 因此十分适合用作特征. 步态能量图 (gait energy image, GEI) 是一种经典的基于轮廓的步态特征, 这是一种平均轮廓模板 [3]. 基于它的改进有步态熵图 (gait entropy image, GEnI) [4] 以及掩码步态能量图 (masked GEI) [5]. 这些改进主要目的是降低衣物和携带物品等因素的影响.

目前, 基于深度神经网络 (deep neural network, DNN) 的技术在图像处理领域取得了巨大的成功, 例如基于深度学习的行为识别 [6], 人脸识别 [7] 等. 在步态分析领域, 深度学习方法也得到了应用 [8, 9]. 卷积神经网络很好地利用了图像数据的特点, 可以自动逐层抽取图像的特征, 在精度方面是当前最好的方法. 但是, 深度神经网络通过大量线性和非线性的组合构造复杂的可学习函数, 这导致人们难以解释它决策的依据. 另外, 深度神经网络需要训练大量的参数, 这导致模型规模大, 需要大量的训练数据, 当问题规模较小时, 较少的训练数据难以训练深度神经网络.

数据挖掘的任务不仅仅是预测新的样本, 我们还希望发现数据中的知识, 使人理解问题的关键所在. 在实际问题中, 也往往需要解释算法决策的依据. 在目前的很多实际工作中, 完成工作需要人和人工智能的共同参与, 如果人能够理解识别算法的根据, 则有利于实际工作的展开. 例如, 在技术人员调试模型时, 可解释性可以提供解决问题的线索; 在安防应用中, 侦查人员可以针对监控系统挖掘到的目标特点进行重点排查; 在智能家居应用中, 理解机器的决策依据有利于企业提供个性化的服务以及进行针对性的营销. 因此, 探索具有可解释性的步态识别算法具有深刻的现实意义. 上文提到, 深度神经网络内在的属性决定了这种算法是一个黑盒, 难以解释决策依据. 目前在步态识别领域的文献中, 关注可解释性的研究尚处于空白状态.

从现实的需求出发, 我们认为解释步态识别的基本要求是指出每一个个体的辨别性步态特点. 具体而言, 我们需要指出什么步态特征在视频中哪几帧的表现是最能辨识这个个体的特征. 这也是本文算法在可解释性层面要达到的目标.

为了提供这种可解释性, 我们用兼具辨别性和可解释性的特征来表示步态. 很多基于模型的方法将步态特征 (例如角度、步长等) 随时间变化的情况表示为时间序列 [10~13], 但这些方法多以简单的最近邻 (1-nearest neighborhood, 1-NN) 方法进行识别. 很多时间序列分类领域的算法可以借鉴到这一问题中. 基于 Shapelet 的方法是一类基于树模型的时间序列挖掘方法. 决策树和随机森林是经典的具有较好可解释性的机器学习算法, Shapelet 树和森林可以看作经典算法在时间序列分类问题上的推广, 因此十分自然地继承了可解释的优良特性. Shapelet 是时间序列中最具辨别性的, 阶段独立的子序列 [14]. 所谓辨别性, 是指子序列与某一类的时间序列的某些子序列相似, 与其他类的任何时间序列的子序列都不相似的性质, 因而可以使用这种序列区分不同类的数据. 由于 Shapelet 本身是某一类特有的特征, 它指出了分类过程的关键点, 这是这种方法提供可解释性的根本原因. 最早的 Shapelet 决策树有准确率较低、训练很慢等缺点. 为了解决这些问题, Karlsson 等 [15] 提出了 Shapelet 森林, 本文作者提出了组合 Shapelets 森林 (random pairwise Shapelets forest, RPSF) [16], 这两种方法很好地提升了算法的分类准确性, 降低了训练模型的开销, 又保留了原算法良好可解释性的优点.

基于以上的讨论, 本文将步态识别抽象为时间序列分类问题, 提出一种基于时间序列表示和组合 Shapelets 森林的步态识别方法. 首先进行人体姿势估计, 将人的拓扑结构表示为棍状模型. 之后从得到的拓扑结构中抽取多种特征, 包括步长、各关节的角度特征等, 并将这些特征随时间变化的情况表

示为时间序列. 最后, 我们将组合 Shapelets 森林推广到多维时间序列分类问题中, 并用其分析上一步得到的特征序列, 达到识别的目的. 在 CASIA-B 数据集上的实验表明, 这种方法在较小规模的问题上具备接近深度学习方法的准确率, 同时还能够对得到的模型进行细粒度的解释, 具体到步态视频的每一帧下, 说明决策的依据. 另外, 这种方法的训练速度也比深度学习方法快.

本文剩余部分组织如下: 第 2 节阐述领域内的相关工作. 第 3 节阐述步态特征提取过程. 第 4 节将组合 Shapelets 森林推广到多维时间序列分类问题上, 并说明如何解释该模型. 第 5 节进行实验验证提出方法的有效性. 第 6 节总结全文.

2 研究背景

2.1 步态识别和卷积神经网络

关于步态识别的研究始于心理学实验 [17], 研究人员在实验中发现不同的人有其特有的行走习惯和姿势. 1973 年, Johansson [18] 的研究工作第一次证实了步态可以作为个体身份的辨识标志. 随着计算机视觉技术的不算发展, 步态识别作为一种远距离生物识别技术发展起来.

用来表示步态的特征对于步态识别性能有重要影响. 步态特征主要分为基于模型和非模型的方法. 人们对基于模型的方法进行了充分的研究. Cunado 等 [19] 最初提出基于下肢关节点的钟摆模型, 并用旋转角度、相位、频率等作为特征进行识别. Fujiyoshi 等 [20] 则分析轮廓点到质心的距离, 采用该距离曲线的极值构建星形模型, 从而分析步态. 需要指出的是, 很多步态特征都具有类似时间序列的特性. 例如, Kale 等 [10] 使用步态宽度和二值轮廓作为特征, 使用隐马尔科夫 (Markov) 模型进行识别, 并且在分析中发现宽度特征具有良好的辨别性. Rustagi 等 [21] 用轮廓质心坐标作为特征. Shajina 等 [11] 使用时间序列 Shapelet 进行步态识别. Wang [12] 和 Yam 等 [13] 根据人体关节角度等特征的运动情况构建特征向量. 这些特征都是具有一维逻辑关系的实值序列, 可以被看作是时间序列.

步态能量图是一种经典的非模型步态特征 [3]. 这是步态轮廓图像在一个步态周期内的平均. 由于这种表示的简洁性和有效性, 步态能量图被广泛地在各种方法中使用. Bashir 等 [4] 提出步态熵图来改进步态能量图, 这种表示通过计算步态轮廓的熵得到. Iwama 等 [22] 评估了多种非模型步态表示在识别任务中的性能, 发现经典的步态能量图仍然表现最好. Lishani 等 [23] 结合步态能量图不同区域的特征和该区域的纹理特征进行步态识别.

卷积神经网络 (convolutional neural network, CNN) 近年来在图像处理领域取得了巨大成功. 这种方法被用来处理各种各样历史悠久的计算机视觉任务, 并且势如破竹地打破已有记录. CNN 首先在图像分类任务上崭露头角. Krizhevsky 等 [24] 在 2012 年训练了一个有 5 个卷积层和 3 个池化层、在当时被认为很深的卷积神经网络, 在 1000 类别图像分类任务中极大地提升了最佳准确率记录. 2014 年, 受到广泛关注的 GoogLeNet 又在该任务中将 top5 的错误率显著降低 [25]. 在图像目标检测任务中, 基于 CNN 的方法也渐渐统治了该领域. Girshick 等 [26] 在 2014 年提出了一种用 Selective Search 搜索候选区域, 用 CNN 提取特征, 最后用支持向量机和线性回归器进行目标区域分类和调整的目标检测框架, 称作 RCNN, 并且迅速流行开来. 很多后续的工作 [27, 28] 进一步改进了 RCNN. 除了 RCNN 系列算法, 其他效果良好的目标检测算法也是基于 CNN 的. 例如 He 等 [29] 提出的 SPPNet, 使用空间金字塔改进了池化层, 达到了良好的目标检测效果. CNN 的研究甚至使一些无人问津的研究领域变得具有价值, 例如图像风格迁移和图像生成等 [30~32].

在人类身份识别的任务上, CNN 被广泛用于各种生物特征. 例如 2014 年, Taigman 等 [7] 在人脸

识别上达到了和人类相等的表现. 他们使用一个超大的人脸数据集训练了一个身份分类网络, 这个数据集包含了来自四千余个个体的 440 万个标记面孔. 对于包含在数据集之内的个体, 可以直接使用这个网络进行识别. 对于更开放的任务, 则使用该网络初始化一个暹罗网络, 使用该网络进行人脸相似性的比对.

CNN 也被应用在步态识别任务中. 目前基于 CNN 的步态识别方法可以分为两种, 第 1 种和经典机器学习问题的思路一致, 直接用画廊集训练 CNN, 对探针集中的步态特征进行分类, 以此达到识别目的. 但问题在于, 步态识别问题是很开放的问题, 在训练 CNN 时人们往往没有识别目标的数据, 考虑到训练 CNN 耗时耗力耗材, 这使得这种方法的应用受到了限制. 第 2 种则使用不属于画廊集和探针集的额外的步态数据训练特征提取模型, 再用该模型提取画廊集和探针集的特征, 最后使用近邻法进行识别. 这种方式更加复杂, 但不要求识别对象在训练 CNN 的数据中, 训练一个 CNN 模型可以识别广泛的群体, 适用于开放的步态识别任务. Alotaibi 等 [8] 提出了一种定制的卷积神经网络, 通过一对一的特征图连接降低了模型的参数规模, 在 CASIA-B 数据集上取得了很好的效果. Shiraga 等 [9] 则提出了 GEINet, 一种以步态能量图为输入的深度卷积网络, 并且在 OU-ISIR 数据集上取得了很好的准确率. Feng 等 [33] 先用 CNN 提取关节热图 (joint heat map), 再将其输入 LSTM 网络进行识别, 以充分利用时序信息. Wu 等 [34] 使用卷积层和加法层模拟加权的 GEI 减法, 再使用卷积网络进行相似度鉴别, 再用该相似度作为距离度量进行最近邻分类. 他们探索了在不同时机进行减法的差异, 还尝试了展开空间信息以代替 GEI. Chao 等 [35] 将步态序列中的轮廓看作一个集合, 提出了 GaitSet. 这种方法不需要输入具有连续性, 使得应用于近邻法的特征抽取器有了更好的灵活性. Zhang 等 [36] 则设计了一种暹罗网络进行基于最近邻的步态识别. McLaughlin 等 [37] 则提出了一种结合了卷积网络和循环网络的新网络训练步态视频嵌入, 并且用于步态重识别任务. 另外, Yu 等 [38] 还探索了用生成对抗网络提取步态特征的方法.

2.2 基于 Shapelet 的时间序列分类

Shapelet 是时间序列中最具辨别性的, 阶段独立的子序列 [14]. 所谓辨别性, 是指子序列与某一类时间序列的某些子序列相似, 与其他类时间序列的任何子序列都不相似的性质, 因而可以使用这种序列区分不同类的数据. 基于 Shapelet 的方法有以下特点. 第一, Shapelet 是一种局部特征, 这也是它和近邻法的主要区别. 第二, 因为在分类时只需要和 Shapelet 作比较, 它在分类阶段快速, 需要较少的存储空间. 第三, 由于 Shapelet 本身是某一类特有的特征, 它指出了分类过程的关键点, 能够提供很好的可解释性. 在初步的工作中, 人们抽取 Shapelet 并嵌入决策树 [14]. 但这种方式的准确率和训练时间都存在较大缺陷. Hills 等 [39] 提出了一种通过一次性遍历数据集提取 Shapelet, 再基于 Shapelet 将时间序列数据集做转换的方法. 这种方法分离了 Shapelet 提取和分类的过程, 使得多种成熟的分类技术得以应用. Karlsson 等 [15] 则引入随机森林, 通过集成一系列训练成本低廉的分类器得到较高的准确率. 本文作者 [16] 则提出组合 Shapelets 森林, 通过在树节点中组合来自不同类的 Shapelets, 进一步提升了 Shapelet 森林的训练速度、准确率以及可解释性.

3 步态特征抽取

本文的方法分为两步, 步态特征抽取和分类识别. 其中, 步态特征抽取是指将步态视频转换为步态特征时间序列的过程. 本节详细地描述这一过程.

图 1 (网络版彩图) 选取轮廓边界用于回归腿部倾斜角度. 图中红色的轮廓是大腿轮廓, 绿色的轮廓是小腿轮廓

Figure 1 (Color online) Selecting contour boundary to estimate the leg angle. The red line denotes the outline of the thigh, and the green line denotes the outline of the shin

3.1 人体拓扑结构估计

首先, 使用背景减除法提取步态轮廓, 该方法在大量文献中均有描述 [3, 13].

之后, 为了抽取步态特征, 首先根据轮廓图像对人体的拓扑结构进行建模. 这里我们使用棍状模型近似人体的结构, 也就是把头部、躯干、大腿、小腿这些人体主要部分表示为一条线, 共计六条线. 把主要关节表示为线的交点. 来自解剖学的人体比例信息 (可参考文献 [40] 中的图 3) 可以作为估计关节点位置的依据 [40]. 由于人行走的过程中下肢运动较为复杂, 我们还需要一种更精密的方法来估计下肢各部分的具体位置. 这种方法分为两步, 首先通过线性回归估计大腿和小腿的倾斜角度, 再对下肢各个关节的位置进行估计. 本节具体介绍估计人体拓扑结构的方式.

3.1.1 腿部倾斜角度估计

我们使用边界像素的线性回归进行大腿和小腿的倾斜角度的估计. 假设人的步行方向向左. 首先根据人体比例信息粗略地估计出骨盆、膝盖和脚踝的纵坐标位置. 由于膝盖的运动最为复杂, 额外从该纵坐标值周围选择 5 个邻近的纵坐标进行比较, 选择纵坐标对应的轮廓向左侧凸出最为明显的纵坐标作为膝盖纵坐标. 该 “凸出” 的程度可以由 5 个坐标的横坐标确定. 确定几个关节的纵坐标后, 提取从骨盆到左右膝盖的外侧轮廓边界, 以及左右膝盖到左右脚踝的外侧轮廓边界. 大腿和小腿的倾斜角度可以通过对这些边界坐标数据进行一元线性回归得到. 回归斜率即作为腿部倾斜角度的估计. 图 1 形象地描述了这一过程.

式 (1) 是一元线性回归中回归斜率的计算公式, $\Theta_{\text{jnt},k}$ 表示第 k 帧大腿或小腿的回归斜率, $\text{jnt} \in \{l_{\text{thigh}}, l_{\text{shin}}, r_{\text{thigh}}, r_{\text{shin}}\}$.

$$\Theta_{\text{jnt},k} = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}, \tag{1}$$

其中 n 表示这一段轮廓边界的像素点数量, x_i, y_i 表示轮廓边界中第 i 个像素点的横纵坐标, $\overline{x}$, $\overline{y}$ 则表示该段轮廓边界的横纵坐标均值. $\Theta_{\text{jnt},k}$ 就是估计得到的第 k 帧中大腿或小腿的倾斜度. 后文将使用这个倾斜度对下肢关节位置进行估计.

在收集用于回归的轮廓坐标数据的过程中, 为了去除噪声干扰, 在自上到下收集大腿边界数据时, 如果相邻纵坐标的点横坐标差距过大, 我们就抛弃已有的点, 这是因为已有的点往往是误收集的手臂数据. 在收集小腿边界数据时, 如果出现相邻纵坐标的点横坐标差距过大, 我们就不再继续收集点, 这

是因为这种情况往往是收集到鞋子的边界数据导致的.

另外, 我们还使用一种分段的回归方式. 首先将用于回归的数据根据纵坐标等分为上下两段, 分别回归上段、下段和全部数据. 若上下两段的斜率异号, 这说明数据的收集出现了噪声, 我们使用上段的斜率作为最终结果. 这是因为在这一步收集大腿数据时, 往往因为错误地收集到了小腿的数据而产生噪声, 在收集小腿数据时, 往往因为错误地收集到了鞋子数据而产生噪声. 因此上段的斜率是更可靠的. 如果上下两段斜率同号, 我们则使用全部数据回归得到的斜率作为估计结果.

3.1.2 关节位置估计

本小节要估计的关节有: 头部, 颈部, 骨盆, 左膝, 右膝, 左踝, 右踝. 上肢在步态运动中比较稳定, 因此头部、颈部、骨盆的纵坐标可以通过高度比例得到 (可参考文献 [40] 中的图 3). 头部、颈部、骨盆的横坐标则可以通过式 (2) 计算得到.

$$x_{\text{center}} = x_l + (x_r - x_l) \div 2, \tag{2}$$

其中 x_l 和 x_r 是对应行上的第一个和最后一个前景像素点. 式 (2) 即寻找中心点作为横坐标. 下肢在步态运动中的状态比较复杂, 估计算法也更精细. 首先通过 3.1.1 小节的方法估计得到左、右、大、小腿的倾斜角度, 之后以骨盆坐标为基准, 根据大腿倾斜角度以及使用人体比例信息推得的大腿长度推断膝盖关节坐标. 以左膝为例, 膝盖的坐标估计可以通过式 (4) 和 (5) 得到. 右膝的计算方式是类似的.

$$x_{\text{pelvis-l}} = x_{\text{pelvis}} - \frac{\text{row}(y_{\text{pelvis}})}{4}, \tag{3}$$

$$x_{\text{lknee}} = x_{\text{pelvis-l}} + L_{\text{thign}} \times \cos \arctan \Theta_{\text{lknee}}, \tag{4}$$

$$y_{\text{lknee}} = y_{\text{pelvis}} + L_{\text{thign}} \times \sin \arctan \Theta_{\text{lknee}}, \tag{5}$$

其中 x_{lknee} 和 y_{lknee} 是左膝的坐标, $x_{\text{pelvis-l}}$ 和 y_{pelvis} 是左腿根部基准坐标, 通过骨盆坐标向左略微偏移得到, 这个值可以通过式 (3) 得到, 其中 $\text{row}(y)$ 指纵坐标为 y 的行上的前景像素数量. Θ_{lknee} 是 3.1.1 小节中估计得到的左侧大腿倾斜度, 由于我们讨论的是一帧的信息, 因此在符号中抛弃了代表帧数的下标. L_{thigh} 则是通过人体比例信息估计得到的大腿长度. 类似地, 我们可以根据得到的膝盖坐标估计得到脚踝的坐标, 式 (6) 和 (7) 以左踝为例描述了计算方法. 右踝的计算方式是类似的.

$$x_{\text{lankle}} = x_{\text{lknee}} + L_{\text{shin}} \times \cos \arctan \Theta_{\text{lankle}}, \tag{6}$$

$$y_{\text{lankle}} = y_{\text{lknee}} + L_{\text{shin}} \times \sin \arctan \Theta_{\text{lankle}}, \tag{7}$$

其中 x_{lankle} 和 y_{lankle} 是左脚踝的坐标, x_{lknee} 和 y_{lknee} 是左膝盖的坐标, Θ_{lankle} 是 3.1.1 小节估计得到的左侧小腿倾斜度, 这里也抛弃了帧数的下标. L_{shin} 是通过人体比例信息估计得到的小腿长度. 到此得到了包含 7 个关节位置的坐标集合 $\{x_{\text{part}}, y_{\text{part}} | \text{part} \in \{\text{head}, \text{neck}, \text{pelvis}, \text{lknee}, \text{rknee}, \text{lankle}, \text{rankle}\}\}$, 也就完成了人体姿态估计. 根据这些坐标, 我们可以得到步态特征序列.

3.2 步态特征序列提取

首先抽取以下特征: 左膝角度 Θ_{lknee}; 右膝角度 Θ_{rknee}; 左脚踝角度 Θ_{lankle}; 右脚踝角度 Θ_{rankle}. 这是因为角度等特征可以反映步态的动态信息, 而动态特征主要集中于下半身. 另外我们还抽取头部

和肩部连线的角度 Θ_{neck}. 上文提到的角度是关节倾斜方向和地面垂线的夹角. 角度可以根据 3.1 小节得到的关节坐标, 通过式 (8) 估计得到.

$$\theta = \arctan \frac{x - x'}{y - y'}, \tag{8}$$

其中 x, y 是关节坐标的值, x', y' 是前一关节坐标的值, θ 是估计得到的角度.

另外还从每个帧中提取几个辨别性较好的轮廓特征: 轮廓的面积 s, 轮廓的质心纵坐标 c, 轮廓的最大宽度 w, 轮廓的高度 h, 轮廓的宽高比 r. 同一特征在不同帧提取的值可以表示为时间序列. 最终, 我们得到了 10 条步态时间序列, 这些序列可以很好地表示步态视频中所包含的特征.

4 使用多维组合 Shapelets 随机森林进行步态识别

本节详细地阐述如何使用第 3 节抽取的特征序列进行步态识别, 以及如何解释识别结果.

多维时间序列是指, 单个实例中有多个代表特定含义的时间序列, 它们描述实例不同方面的特征, 共同构成一个实例. 这和经典分类问题中的属性或特征的概念类似. 该实例具有一个类标. 多维时间序列分类算法对这样的时间序列数据集构建从序列到类标的映射, 从而对未知类标的实例进行预测. 第 3 节抽取的步态时间序列实例就是典型的多维时间序列实例.

RPSF 的决策树节点中嵌入了来自两个不同类的 Shapelets, 在分类准确率、训练开销、可解释性几个方面都是较为先进的时间序列分类算法. 本节将 RPSF 算法推广至多维时间序列, 提出多维组合 Shapelets 随机森林 (MRPSF), 使其可以直接用于步态识别任务.

由于本节将问题抽象为时间序列分类问题, 因此本节将使用术语训练集/测试集, 而非本文其他部分使用的画廊集/探针集.

4.1 多维组合 Shapelets 随机森林

首先定义多维时间序列分类问题. 给定实例维度相等的、时间序列长度相等的多维时间序列训练集 D, 该数据集的维度数目是 d, 实例个数是 n, 每条时间序列实例的长度是 m. 多维时间序列分类算法从训练集中学习一个从多维时间序列实例到类标的映射 f, 以期在维度、长度和 D 相同的多维时间序列测试集 T 上进行精确的预测.

4.1.1 MRPSF 的参数

首先介绍算法的输入参数. 算法接受几个参数: 决策树的数量 p, Shapelet 长度区间 l, u 和考察 Shapelets 对的数量 r.

下面具体解释算法输入参数的含义. 随机森林由多棵决策树共同构成, p 规定了决策树的数量, 决定了森林的大小. 由于 Shapelet 是时间序列的局部片段, 不宜过短或过长, 因此在决策树节点中搜索 Shapelet 时会限定 Shapelet 的长度, 在本文中这个长度的上下限被设定为参数 l, u. Shapelet 森林中的决策树节点在搜索最佳 Shapelet 时不会搜索全候选空间, 而是从全候选空间随机选择一小部分作为候选. 这是时间序列随机森林中进行特征扰动的具体表现形式. 参数 r 则控制在该上述搜索过程中从全空间中随机抽取的候选集大小. 表 1 列出了 MRPSF 的超参数. 其中 numcand_i 是第 i 个类中可能的 Shapelet 数量, N_c 是类别总数.

在 4.1.2 小节中, 一些子算法的输入中有多维时间序列数据集的维度 d. 但该值由数据集本身决定, 并非一个需要手工设定, 影响算法性能的参数. 因此该值应当被视为输入数据的一部分, 而非算法的

表 1 MRPSF 的参数

Table 1 Parameters of MRPSF

Parameter name	Notation	Value range
Decision tree number	p	$[1, +\infty)$
Shapelet maximum length	u	$(l, m]$
Shapelet minimum length	l	$(0, u)$
Shapelet candidate set size	r	$[1, \sum_{i=1}^{N_c}\sum_{j=1,j\neq i}^{N_c}(\text{numcand}_i \times \text{numcand}_j)]$

参数.

4.1.2 MRPSF 的训练过程

算法 1 展示了 MRPSF 算法的训练步骤, 它为每一个时间序列维度训练一组决策树, 通过集成的方式得到每一维度的最终模型, 并通过集成的方式得到最终的模型. 训练每一棵树时先随机选择多维时间序列中的一个维度, 用 Bootstrap 抽样在该维度中产生训练集 D, 并传入 RandomPairwise-ShapeletsTree 函数构造决策树. 完成训练后需要记住训练数据的来源维度. 如此训练 p 棵树.

算法 1 MultiDimensional-RandomPairwiseShapeletsForest

Input: Multi-dimensional time series dataset $D = \{D_1, \ldots, D_d\}$ (D_i represents the ith dimension time series), dimension number d, Shapelet minimun length l, Shapelet maximum length u, tree number p, Shapelet candidate number r;

for $i = 1$ to d **do**
  $R_i \leftarrow \emptyset$;
end for
for $i = 1$ to p **do**
  dimension=SelectRandomDimension();
  $I_i \leftarrow \text{Sample}(D_{\text{dimension}})$;
  $\text{ST}_i \leftarrow \text{RandomPairwiseShapeletsTree}(I_i, l, u, r)$;
  $R_{\text{dimension}} \leftarrow R_{\text{dimension}} \cup \{\text{ST}_i\}$;
end for
return R;

Output: Set of decision tree sets trained from different dimensions $R = \{R_1, \ldots, R_d\}$ (R_i represents trees trained by the ith dimension).

算法 2 展示了训练每一棵树的过程. 开始时判定是否应该生成叶节点. 生成叶节点的条件是输入数据纯度较高 (熵小于 0.1). 之后从所有符合长度限制的子序列中随机抽取两条来自不同类的子序列构成一个序列对. 准确地说, 随机选择两个类, 再从这两个类的时间序列中随机选择两条时间序列, 之后随机选择两个长度和起始点, 构成一对 Shapelets: s_1 和 s_2. 这样抽取 r 次后, 我们就得到了一个候选 Shapelets 对集合. 然后 BestShapeletsPair 函数使用 AssessCandidatePair 函数 (算法 3) 评判该集合, 从中找到最佳的一对子序列. 之后 Split 函数会将数据集 D 分割. 由于没有分裂阈值, 这个函数将比较 D 中时间序列到 s_1 和 s_2 的距离, 将离 s_1 较近的实例分到一侧, 将离 s_2 较近的实例分到另一侧. 最后在分割好的数据集上递归地训练子树.

在算法 2 中, BestShapeletsPair 函数遍历 S 中的每个候选 Shapelets 对, 并且一一评价. 具体的评价方式由算法 3 描述. 对于一对 Shapelets s_1 和 s_2, 先分别计算它们与训练集中每一条时间序列 $D[m]$ 间的子序列距离 subdist_1, subdist_2. 若 subdist_1 小于等于 subdist_2, 表示 $D[m]$ 与 s_1 的距离较近, 此时 $D[m]$ 被添加到 set_1 中, 否则, $D[m]$ 被添加到 set_2 中. 这一过程结束后, set_1 和 set_2 分别保存了距

算法 2 RandomPairwiseShapeletsTree

Input: Time series dataset D, Shapelet minimum length l, Shapelet maximum length u, Shapelet candidate number r;

if IsTerminal(D) **then**
  return MakeLeaf(D);
end if
$S \leftarrow \emptyset$;
for $i = 1$ to r **do**
  $S \leftarrow S \cup \{\text{SampleShapeletsPair}(D, l, u)\}$;
end for
$(s_1, s_2) \leftarrow \text{BestShapeletsPair}(D, S)$;
$(D_1, D_2) \leftarrow \text{Split}(D, s_1, s_2)$;
$\text{ST}_l \leftarrow \text{RandomPairwiseShapeletsTree}(D_1, l, u, r)$;
$\text{ST}_r \leftarrow \text{RandomPairwiseShapeletsTree}(D_2, l, u, r)$;
return s_1, s_2, ST_l, ST_r;

Output: A decision tree node consisting of a Shapelets pair s_1, s_2, left subtree ST_l and right subtree ST_r. The node can be considered as a decision tree because it contains pointer to subtrees.

算法 3 AssessCandidatePair

Input: A Shapelets pair s_1, s_2, time series dataset D;

$\text{Gain} \leftarrow 0$, $\text{Gap} \leftarrow 0$, $\text{line}_1 \leftarrow 0$, $\text{line}_2 \leftarrow 0$;
for $m = 1$ to $|D|$ **do**
  $d_1 \leftarrow \text{subdist}(s_1, D[m])$;
  $d_2 \leftarrow \text{subdist}(s_2, D[m])$;
  if $d_1 \leqslant d_2$ **then**
    $\text{set}_1 \leftarrow \text{set}_1 \cup D[m]$;
  else
    $\text{set}_2 \leftarrow \text{set}_2 \cup D[m]$;
  end if
end for
$\text{Gain} \leftarrow \text{InfoGain}(\text{set}_1, \text{set}_2)$;
$\text{Gap} \leftarrow \text{SepGap}(\text{set}_1, \text{set}_2)$;
return Gain, Gap;

Output: Information gain Gain, separation gap Gap.

离 s_1 或 s_2 较近的实例, 两者之间没有交集. 这相当于完成了该决策树节点的数据划分. 信息增益和分割间隙将被用来评判该分割的好坏 [15], 并且以此作为 Shapelets 对质量的度量. 信息增益是决策树算法中度量数据集类标纯度的常用方法, 启发式决策树希望分裂得到的两个数据集尽可能纯净. 分割间隔则是当多个 Shapelets 对产生相等的信息增益时启动的评判机制, 这种机制基于类似支持向量机的思想, 希望尽可能将两侧数据的距离拉大. 这两个评分最终会返回给 BestShapeletsPair 函数.

训练完成后, 算法的分类过程十分简单. 决策树的每一个内部节点都由序列对 (s_1, s_2), 以及该节点的左右子树构成. 而叶节点则记录了类标值. 分类时从根节点开始计算待分类实例和两个 Shapelets 的距离, 若测试实例 T 与 s_1 间的子序列距离小于它与 s_2 间的子序列距离, 则递归地调用左子树进行分类. 否则, 将递归地调用右子树进行分类. 该过程不断重复, 直至到达叶子节点并得到预测类标. 我们有用来自各个维度时间序列训练的总共 p 棵树, 对于每棵树都用测试实例中对应维度的时间序列进行这一过程并进行投票, 最终返回得票最多的类标作为预测类标.

4.2 使用步态 Shapelet 森林解释个体步态特点

本文算法的优势是, 可以较为具体地指出个体的步态特点. 具体而言, 我们的算法可以指出哪些特征在哪几帧的表现对某个个体而言最具有辨别性和代表性. 本小节详细介绍解释步态识别模型的方法. 我们首先介绍分解不纯度下降, 这是组合 Shapelet 森林的特征重要度度量方法. 再使用该方法解释步态识别问题.

4.2.1 分解不纯度下降

基于 Shapelet 的决策树的优势之一是具备可解释性. 尽管引入随机化和集成使得算法性能有很大提升, 但人们在解释模型时会遇到困惑. 不同树的解释存在不同甚至矛盾. 在经典随机森林中同样存在这种问题. 一种解决方式是, 将森林中所有分裂属性为某一相同属性的节点的信息增益加起来, 构成这个属性在森林中重要度的评分 [41]. 这种方法称作平均不纯度下降 (mean decrease impurity, MDI) 度量.

对于 Shapelet 森林, Karlsson [15] 将类似的度量方式进行了推广. 他在决策树每个节点中记录得到信息增益, 并把该增益加到构成该节点 Shapelet 的时间步上面, 最终得到时间步的评分. 基于类似的思想, 我们也为组合 Shapelet 森林定义了评分策略. 和文献 [15] 中方法的主要不同点和优势是, 我们为每个类都定义了时间步评分, 可以知道任何一个类的辨别性模式. 这种优势归功于 RPSF 的树结构. 对于任何一个节点, 我们将嵌入的一对 Shapelets 产生的信息增益分解, 以辨别具体是哪个 Shapelet 产生了较大的贡献. 分解的原则是, 对于一个 Shapelets 对 (s_1, s_2), 如果 s_1 吸引的时间序列构成的数据集造成了更大的不纯度降低, 那么就认为 s_1 为系统贡献较多. 之后我们会把 Shapelet 的贡献加到构成 Shapelet 的那段时间步上, 不同类的 Shapelet 会分别累计评分. 基于以上的想法, 我们定义分解平均不纯度下降 DMDI. 给定单维组合 Shapelets 森林 $R = \mathrm{ST}_1, \mathrm{ST}_2, \ldots, \mathrm{ST}_n$, 其中 ST 是组合 Shapelets 决策树, 其中有多个树节点 node, 每个节点对应着 Shapelets 对 pair $= (s_1, s_2)$. 又给定等长时间序列训练集 D, 序列长度为 m. 那么对于 D 的每一个时间序列属性 k 和每一个类 c, 定义分解平均不纯度下降 DMDI(k, c) 为

$$\mathrm{DMDI}(k,c) = \sum_{p}\left(\sum_{\mathrm{node}\ (k\in S_1 \wedge \mathrm{class}(S_1)=c)} \mathrm{CV}(\mathrm{node}, S_1) + \sum_{\mathrm{node}\ (k\in S_2 \wedge \mathrm{class}(S_2)=c)} \mathrm{CV}(\mathrm{node}, S_2)\right), \tag{9}$$

其中 CV 是某一条 Shapelet 为节点信息增益做出的贡献值, 这由分解节点的总信息增益得到. 设输入到 node 的数据集是 D_0, node 分割 D_0 后得到的数据集是 D_1 和 D_2, $I(s_1, s_2)(D_0)$ 是该 Shapelets 对的信息增益 [15], 则

$$\mathrm{CV}(\mathrm{node}, s_i) = \frac{\mathrm{ER}(\mathrm{node}, s_i)}{\mathrm{ER}(\mathrm{node}, s_1) + \mathrm{ER}(\mathrm{node}, s_2)} \times I(s_1, s_2)(D_0), \tag{10}$$

$$\mathrm{ER}(\mathrm{node}, s_i) = E(D_0) - E(D_i), \tag{11}$$

其中 ER 是某一条 Shapelet 造成的不纯度降低. 我们并不能保证 ER 得到正值, 如果得到负值, 则将 ER 置为零. 若分母两项同时为负, 则抛弃这个 node.

重新定义的 DMDI 为一个类找出所有嵌入了来自这个类的 Shapelet 的节点. 分解这些节点的信息增益, 把来自这个类的 Shapelet 应得的部分加到该 Shapelet 对应的序列属性上, 最终我们可以知道序列中的哪些属性为特定的类贡献了较大的信息量.

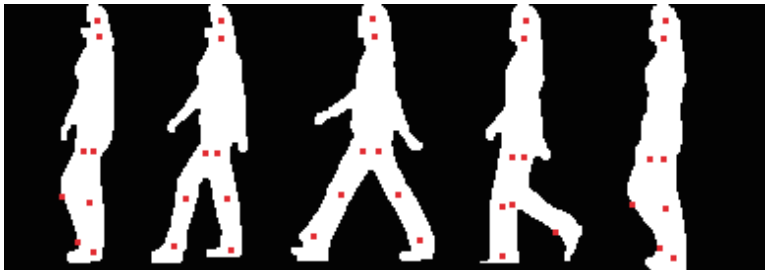

图 2 (网络版彩图) 人体姿态估计效果

Figure 2 (Color online) Results of human pose estimation

4.2.2 解释步态识别

第 1 节提到, 从现实的需求出发, 我们认为解释步态识别的基本目标是, 指出每个个体的辨别性步态特征.

我们从特征序列维度和视频帧两个层面对步态特征的重要性进行评价. 获得该重要性评分后, 我们就可以知道某个个体的步态和其他个体最显著的区别在哪里, 达到我们提出的目标.

首先, 从特征序列维度的层面对几个步态特征进行评价. 由于每个特征都有一个子森林. 我们使用这些子森林在验证集上测试. 该验证集准确率可以作为特征辨别性的评分.

之后, 针对某一辨别性较好的具体特征, 分析视频每一帧的辨别性. 某一具体特征的森林是单维度的组合 Shapelets 森林, 因此可以直接对其使用 DMDI 进行序列重要度分析. 进而可以知道哪些特征在视频的哪几帧对特定个体具有较好辨别性. 这是对具体问题非常详细的解释.

5 实验与分析

本节通过实验说明本文提出的方法的有效性. 我们在 CASIA-B 数据集上展开了一系列实验, 和几种较新的卷积神经网络进行对比, 从识别准确率、训练时间消耗、可解释性 3 个角度评估 MRPSF, 最终说明我们的算法有较好的综合表现. 我们还评估了部分算法参数对算法性能的影响. 另外, 在识别准确率部分, 我们还讨论了几种较新的基于卷积神经网络的步态识别方法在较小规模问题上的性能及原因, 相关文献中尚未有类似实验.

5.1 实验环境和数据集

实验的软硬件环境如下. 软件上, 我们分两个模块在 Windows10 上实现了我们提出的算法. 特征提取部分使用 Python3.5, 基于 OpenCV3.4 实现. MRPSF 则使用 Java1.8 实现. 我们在 Ubuntu16.04 上使用了 CUDA7.0 和 CUDNN9.2 加速神经网络的训练. 硬件上, 我们使用了 Intel 四核 i7 6700 处理器进行我们提出的算法的实验工作. 另外, 我们使用一台搭载 Intel 四核 i7 6700 处理器和 NVIDIA GTX750Ti GPU (2G VRAM) 的计算机进行用于对比的神经网络的训练.

CASIA-B 数据集由中国科学院生物识别与安全技术研究中心建立, 是一个大规模、多角度的步态数据库[1)]. 该数据库共包含 124 人, 每个人有从 11 个视角捕捉、在 3 个状态下行走的视频片段.

首先, 图 2 展示了一些人体姿势估计的效果, 可以看出本文的方法可以有效地对人体的拓扑结构建模.

1) http://www.cbsr.ia.ac.cn/english/Gait%20Databases.asp.

5.2 对比算法

One-on-one Net [8]: 这是文献 [8] 中提出的使用一对一连接卷积层 (one-on-one connection convolution layer) 的特殊卷积神经网络, 原文并未对该网络命名, 为了方便, 本文称之为 One-on-one Net. One-on-one Net 在 CASIA-B 数据集上进行全识别可以达到 98.3% 的准确率. 就准确率而言, 是一种在该数据集上十分成熟的方法.

GEINet [9]: 这是一种八层的卷积神经网络, 包含两个卷积层 – 池化层 – 规范化层三元组, 以及两个全连接层.

GaitSet [35]: GaitSet 是一种先进的步态特征抽取方法. GaitSet 以集合的角度看待训练数据, 使其不需要具备连续性. 这种特性使得这种算法的训练具有很好的灵活性. 这种方法和前两种方法略有不同. 它先使用一些数据训练特征提取模型. 在进行步态识别时, 使用训练好的模型分别抽取画廊集和探针集的特征, 再通过最近邻进行识别.

LBNet [34]: LBNet 是一种使用卷积神经网络计算 GEI 相似度的方法. 它使用卷积层和加法层模拟 GEI 的加权减法, 再使用多层神经网络提取步态特征. 和 GaitSet 类似, 它首先使用一些步态数据训练相似度计算模型, 在进行步态识别时, 以该相似度作为距离度量, 用最近邻进行识别.

5.3 识别准确率

本小节在识别准确率层面, 在较小的问题规模下评估了我们提出的方法和文献中的深度学习方法.

5.3.1 实验设置

对于每一种行走姿势, 我们都在 CASIA-B 数据集中分别选取 5 个人, 8 个人, 12 个人和 15 个人进行识别实验. 并且记录识别的平均准确率.

MRPSF: 我们的方法基于从侧面拍摄的行走视频, 因此选取数据集中的 90 度行走状态的数据进行实验. 我们在正常 (nm), 穿大衣 (cl), 背包 (bg) 3 种状态下都进行了实验. 在正常状态下每个个体有 6 个行走视频, 我们将 6 个视频中的 4 个作为画廊集进行学习, 2 个作为探针集进行测试. 在穿大衣和背包两种状态下, 每个个体有两个行走视频, 我们将其中的一个作为画廊集进行训练, 另一个作为探针集进行测试. 由于 MRPSF 算法对 u, l 两个参数比较敏感, 我们对这两个参数进行了网格搜索 (区间为 l: [0.15,0.45] m, u: [0.35,0.6] m, 步长为 0.05 m). 实验结果是 26 组实验的平均值.

One-on-one Net 和 GEINet: 这两种方法训练 CNN 多分类器作为识别器. 对于每个待识别个体, 我们将 4 个正常状态下的步态视频加入画廊集进行学习, 2 个加入探针集进行测试. 在穿大衣和背包两种状态下, 我们将一个视频加入画廊集, 另一个加入探针集. 学习率设置为 0.01, 使用 Adam 优化器, 在画廊集上迭代至收敛. 实验结果是 26 组实验的平均值.

LBNet: 这种方法的训练过程和上述的方法不同. 给定问题规模 m, 我们先选取 m 个人的所有步态数据训练特征提取器. 之后再选取 m 个人, 分别将 4/1/1 条正常/穿大衣/背包行走视频作为画廊集, 剩下的 2/1/1 条正常/穿大衣/背包行走视频作为探针集, 用近邻法识别探针集. 注意实验中不同的行走姿势是没有交叉的. 这和文献 [34] 的实验方案是一致的. 实验结果是 5 组实验的平均值.

GaitSet: 这种方法的训练过程和 LBNet 是一致的. 不同之处在于, 我们在测试时将 4 条正常行走视频作为画廊集, 两条正常行走、两条穿大衣、两条背包视频作为探针集. 这和文献 [35] 采用的实验方案是一致的. 和上面 3 种方法不同, 该实验的画廊/探针集混杂了不同的行走姿势, 这样做的目的在于尽可能保持原文献的识别效果. 实验结果是 5 组实验的平均值.

表 2 正常行走姿势下, MRPSF 和卷积神经网络的准确率对比

Table 2 Accuracy comparison between MRPSF and CNN in normal walking condition

Algorithm	5-person experiment (%)	8-person experiment (%)	12-person experiment (%)	15-person experiment (%)
MRPSF	96.92	97.35	94.55	92.53
One-on-one Net	99.62	97.12	99.04	98.08
GEINet	98.85	98.80	99.52	99.23
GaitSet	100.00	100.00	100.00	100.00
LBNet	94	92.5	98.33	98.67

表 3 穿大衣行走姿势下, MRPSF 和卷积神经网络的准确率对比

Table 3 Accuracy comparison between MRPSF and CNN in clothing walking condition

Algorithm	5-person experiment (%)	8-person experiment (%)	12-person experiment (%)	15-person experiment (%)
MRPSF	97.69	93.75	89.42	88.97
One-on-one Net	99.23	96.63	94.87	97.18
GEINet	100.00	97.12	98.08	97.69
GaitSet	86.00	87.50	73.33	88.67
LBNet	88.00	75.00	86.68	84.00

表 4 背包行走姿势下, MRPSF 和卷积神经网络的准确率对比

Table 4 Accuracy comparison between MRPSF and CNN in bagging walking condition

Algorithm	5-person experiment (%)	8-person experiment (%)	12-person experiment (%)	15-person experiment (%)
MRPSF	96.15	92.79	92.63	89.23
One-on-one Net	94.62	95.19	91.99	92.82
GEINet	95.38	95.19	93.91	94.36
GaitSet	100.00	100.00	97.50	100.00
LBNet	96.00	67.5	88.33	93.33

5.3.2 实验结果和讨论

表 2~4 展示了本实验的结果.

通过本实验可以看出, MRPSF 在多种情况下都得到了 90% 左右或更高的准确率. 这说明了这种方法在识别率上的有效性. 然而和其他的方法相比, MRPSF 在很多任务上表现出一定的差距. 例如, RPSF 全面差于 One-on-one Net 和 GEINet, 在正常状态下, 也差于 GaitSet. 我们认为原因是, 要识别多个不同的个体, 需要非常细致的特征. CNN 方法的输入是步态能量图, 该步态特征能够描述大量的细节信息, 深度神经网络还可以逐层自动学习特征. 而为了设计具有可解释性的算法, 我们手工设计了时间序列特征作为 MRPSF 的输入, 在转化过程中造成了一定的信息损失, 也不能保证提取的特征是对识别任务而言最好的特征. 因此, 在以后的工作中, 可以尝试研究更好的特征, 增加更多的特征, 以提升算法的识别性能.

我们观察到特别的实验现象: LBNet 表现相对较差, 和在原始文献中的表现不相符. GaitSet 在正常和背包状态下较好, 但在穿大衣状态下较差. 该现象在提出 GaitSet 的文献中并未记述. 我们认为原因如下. 深度神经网络需要大量数据拟合大量参数. 同时这两种方法有共同点: 它们训练 CNN 特征

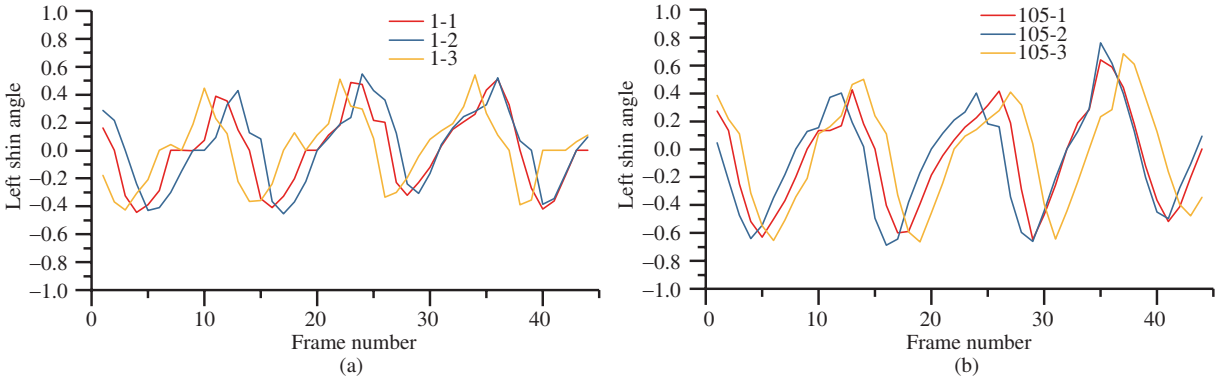

图 3 (网络版彩图) (a) 1 号和 (b) 105 号个体的左侧小腿角度时间序列示意

Figure 3 (Color online) Time series of subject (a) No. 1 and (b) No. 105's left shin angle

抽取模型, 再用这个模型抽取其他图像数据的特征进行近邻分类. 测试阶段使用的近邻法和训练模型时采用的监督学习方法是不同的, 这一点不同对特征质量提出了更高的要求, 同时要适用于 CNN 训练数据中未见的个体也要求提取的特征要具有更好的泛化能力. 上述两点特点使得这种方法对数据量的要求较高. 在原文献中, 作者使用了大量数据训练该特征提取模型, 得到了很好的泛化效果. 而具体到本文研究的小规模问题, 则没有足够的数据支持该泛化, 导致这两种方法的特征提取模型产生了过拟合, 未能提取高质量的特征. 这也说明 RPSF 能够弥补小规模问题下基于 CNN 的方法的缺陷.

5.4 实例分析

MRPSF 对于步态识别问题可以进行详细、细致的解释. 本小节观察数据库中个体的差异, 并设计一些实验检验算法能否发现这些差异, 以说明提出的解释策略的有效性.

由于 Shapelet 是阶段独立的, 当 MRPSF 应用于分类问题时, 对于时间序列的对齐并没有要求, 即不要求同一时间步对应步态周期的同一状态. 但当我们使用 DMDI 对模型进行解释, 则希望时间序列在逻辑上尽量对齐. 这是因为 DMDI 在累加信息增益时将不同时间序列的相同位置同等看待. 因此, 本实验根据宽度序列对步态周期进行了对齐.

首先, 我们根据单独维度特征序列的验证集准确率判定特征的重要程度, 然后使用 DMDI 对该维度进行分析以获得具体到帧的解释.

5.4.1 分析两个人的步态差异

本小节实验选取了两个身材类似的男性: 1 号和 105 号, 在这两个人的步态数据上训练二分类模型, 并且使用 DMDI 分析它们的特征曲线.

本实验中, 在使用单维特征时间序列训练的决策树构成的森林进行预测的场景下, 步宽特征序列和左侧小腿倾斜角度特征序列可以达到 100% 的准确率, 这表明这两个特征对于该问题是具有辨别性的特征. 接下来我们在这两维特征上应用 DMDI, 以更加具体地解释模型的决策依据.

首先我们对两者左侧小腿倾斜角度的特征时间序列进行分析. 图 3 展示了 1 号个体和 105 号个体的左侧小腿倾斜角度序列. 图 4 则展示了在该维度应用 DMDI 统计方法的结果.

从图 3 可以看出, 两类曲线的峰值和低谷有较大差别. 105 号个体的峰值和低谷的绝对值明显比 1 号个体大, 这表现了二者步行过程中小腿摆动最高点的不同. 反映在图 4 的 DMDI 分析结果上, 1 号个体的 17 帧位置和 105 号个体的 37 帧位置有两个峰值, 分别对应着倾斜角度曲线的一个低谷和一

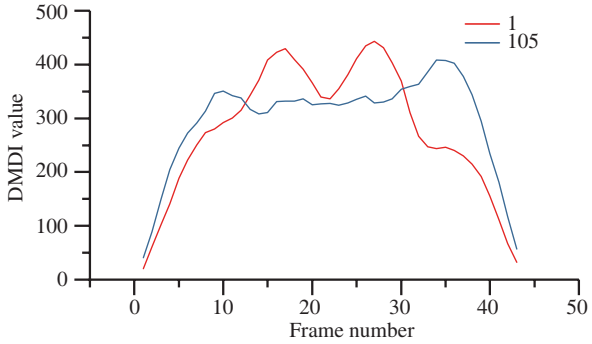

图 4　(网络版彩图) 1 号和 105 号个体的左侧小腿角度时间序列的 DMDI 分析结果

Figure 4　(Color online) DMDI of time series of subject No. 1 and No. 105's left shin angle

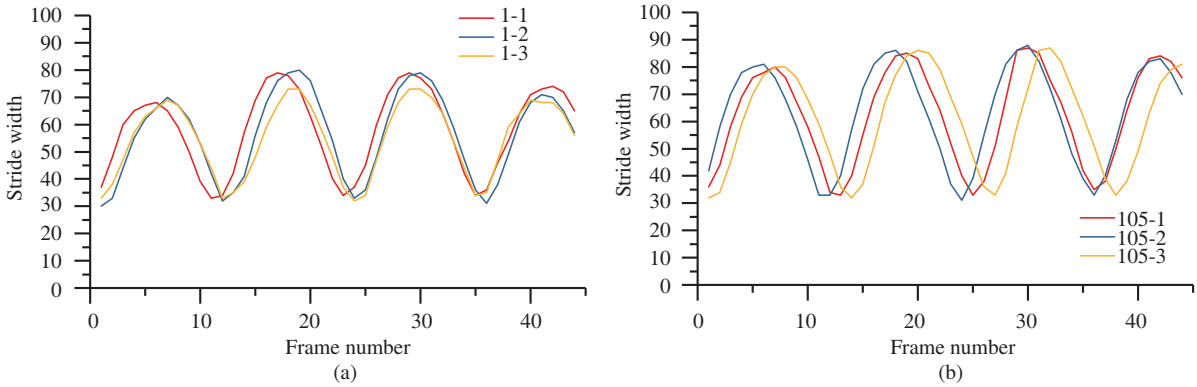

图 5　(网络版彩图) (a) 1 号和 (b) 105 号个体的步宽时间序列示意

Figure 5　(Color online) Time series of subject (a) No. 1 and (b) No. 105's stride width

个峰值.

与此同时, 105 号的序列在下降时明显更加陡峭, 即由于 105 号步宽较大, 导致其腿部摆动速度较快. 图 4 中 27 帧位置的峰值捕捉到了这点差异. 第 10 帧位置出现的峰值用肉眼较难解释. 经过分析算法挖掘的过程, 我们发现 1 号个体的曲线在该位置的之字形爬升具有较好的辨别性, 这是 1 号个体较为独特的步态习惯. 该之字形爬升在第一个步态周期内最明显, 这说明算法捕捉到了非常细微的局部差异.

之后, 我们对两者的步宽序列进行分析. 图 5 展示了 1 号个体和 105 号个体的步宽时间序列. 从图中明显可见 105 号个体的步宽更宽, 表现在序列上即峰值更大, 另外峰值和低谷处的开口也更尖锐, 这对应着 105 号个体在腿部摆动到高点时收腿较快的特点.

图 6 描绘了在步宽序列上应用 DMDI 的结果. 可以看出 1 号个体的 DMDI 曲线呈现较好的周期性, 每一个尖峰都对应着步宽序列的峰值. 这说明算法捕捉到了两者步宽峰值差异较大的特点. 另外, 105 号个体的 DMDI 曲线在大约第 12 和 25 帧的位置出现了峰值, 这对应着两个开口较尖锐的波谷, 该处的子序列具有较好的辨别性. 以上分析表明, 本文提出的方法具有较好的可解释性.

5.4.2　分析 3 个人的步态差异

DMDI 的一个优点是可以对不同的类别提供解释. 5.4.1 小节我们考察了二分类问题, 在二分类问题中, 一个类的辨别性模式反过来也是另一个类的辨别性模式, 因此不能较好体现 DMDI 的优势. 本

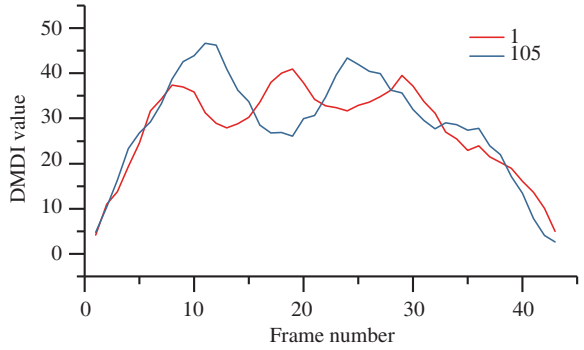

图 6　(网络版彩图) 1 号和 105 号个体的步宽时间序列的 DMDI 分析结果

Figure 6　(Color online) DMDI of time series of subject No. 1 and No. 105's stride width

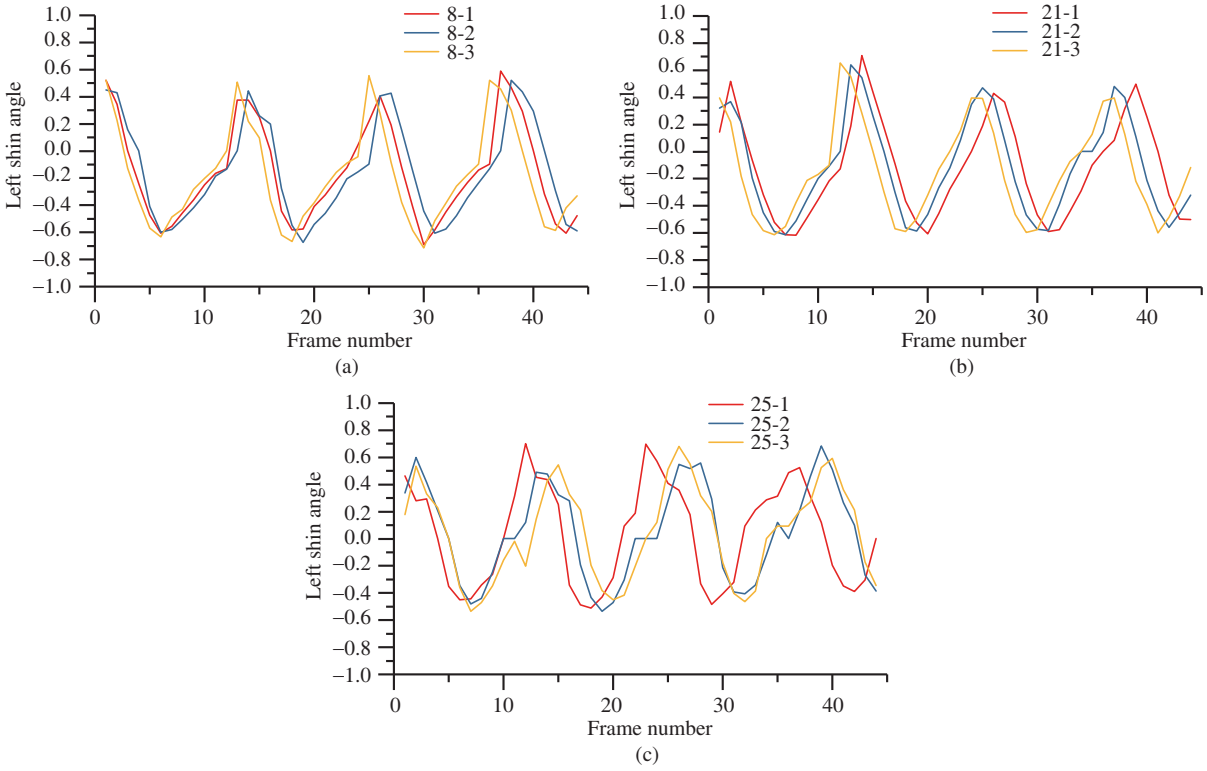

图 7　(网络版彩图) (a) 8 号, (b) 21 号和 (c) 25 号个体的左侧小腿角度时间序列示意

Figure 7　(Color online) Time series of subject (a) No. 8, (b) No. 21 and (c) No. 25's left shin angle

小节的实验选择了 8 号、21 号、25 号 3 个人进行实验.

使用这 3 个人的左侧小腿角度时间序列进行预测可以达到 100% 的准确率, 说明这是一个辨别性特征. 接下来我们分析该结果.

图 7 展示了这 3 个人的小腿轮廓时间序列. 从图中可以看出, 8 号的序列在第 10, 22, 33 帧处的回升较其他人显得缓慢且平缓, 这说明他的腿部摆动较慢. 而 25 号的摆动角度最低点明显比 21 号绝对值小, 21 号的最大摆动角度小于 -0.6, 而 25 号大于 -0.6, 这对应着他们摆动幅度的不同. 这些微妙的区别反映了这几个人走路习惯的细微不同.

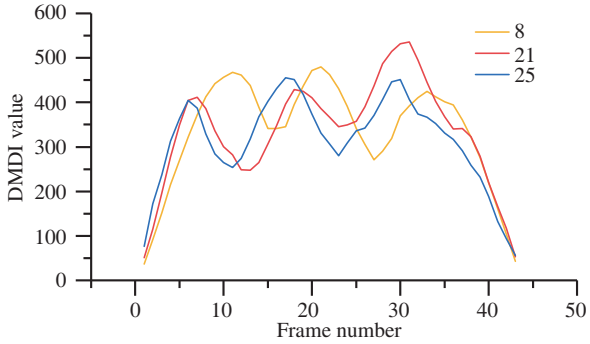

图 8　(网络版彩图) 8 号, 21 号和 25 号个体的左侧小腿角度时间序列的 DMDI 分析结果

Figure 8　(Color online) DMDI of time series of subject No. 8, No. 21 and No. 25's left shin angle

图 8 展示了这个问题的 DMDI 曲线. 可以看出 DMDI 曲线表现出良好的周期性, 契合步态曲线的特点. 在第 10, 22, 33 帧左右的位置, 8 号的 DMDI 曲线达到峰值, 表明算法注意到了此处曲线较为平缓的辨别性模式. 21 号和 25 号的曲线显示出一定的同步变化趋势, 都在第 7, 20, 31 帧左右的位置达到峰值. 这说明算法选择了这里的低谷绝对值差别对这两个类进行区分. 以上分析表明, DMDI 在步态问题上具有较好的可解释性.

5.5　与 CNN 的训练消耗对比

Shapelet 森林的一大优势是只需要考察较少的 Shapelet, 因此需要较少的训练时间. 另外由于本文提出的将步态信息抽取成为时间序列的过程压缩了图像上的步态信息, 使得训练数据规模得到降低, 这也使得本文提出的方法在较短的时间内可以完成训练. 另外, 由于 MRPSF 的森林大小 p 参数对于训练时间有显著的影响, 调节此参数可以进一步缩短训练时间, 尽管这样做会降低准确率, 但仍可以得到可接受的结果. 在具体的应用场景下, 人们可以利用该点优势进行针对性的选择. 本节将 MRPSF 和 CNN 在训练时间的层面进行对比, 以评估 MRPSF 在时间消耗上的表现.

5.5.1　训练时间

本小节设定问题规模为 8 个个体, 在相同的数据集上训练了 MRPSF 和 5.2 小节提到的几种卷积神经网络, 并且给出了它们的时间消耗.

目前, 深度神经网络具有大量的软件和硬件支持来充分利用它的可并行计算特性. 为了公平起见, 在本实验中, 部分实验没有使用 GPU 和 CUDA (compute unified device architecture, 统一计算架构) 等技术, 而是在 CPU 上以串行计算方式对深度神经网络进行训练. 对于明显较慢的算法, 我们使用了 CUDA 加速技术, 否则无法在可接受的时间内得到实验数据. 表 5 的第 3 列给出了对应实验是否使用了 CUDA 加速. 值得一提的是, MRPSF 也具备可并行计算特性. 若对其应用类似的技术, 理论上可将训练时间压缩两个数量级以上.

表 5 给出了实验结果. 第 2 节提到, 目前的深度学习步态识别方法分为两类: 直接分类的方法和抽取特征再分类的方法. 从中可以看出, MRPSF 和属于直接分类方法的 One-on-one Net 差距不大, 但仍比 One-on-one Net 快. 而和属于抽取特征再分类的其他两种方法比较, 则占据绝对的领先地位. 5.3 小节提到, 训练能够抽取具有广泛性的步态特征是比直接分类更困难的任务, GaitSet 和 LBNet 要花费更长时间收敛这一事实也验证了这一点. 实验也表明就训练时间的角度, MRPSF 较基于 CNN 的方法更好. 对于需要进行密集训练的应用, MRPSF 能够满足需求.

表 5 MRPSF 和 CNN 的训练时间对比

Table 5 Training cost comparison between MRPSF and CNN

Algorithm	Training time	Using CUDA
MRPSF	0 h 3 min 59 s	No
One-on-one Net	0 h 5 min 17 s	No
GaitSet	5 h 17 min 26 s	Yes
LBNet	1 h 26 min 17 s	Yes

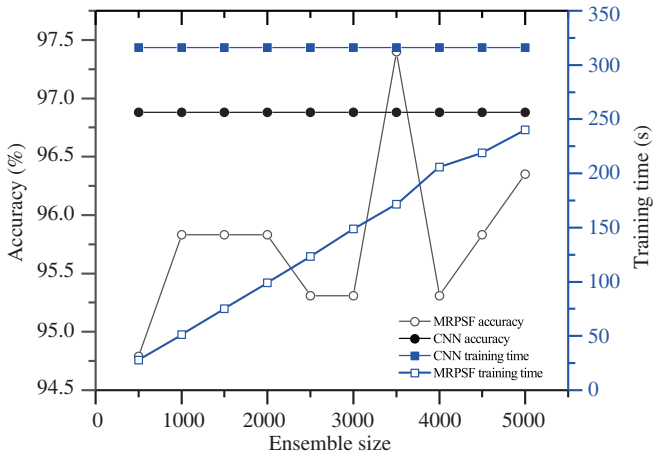

图 9 (网络版彩图) MRPSF 和 CNN 的训练时间对比

Figure 9 (Color online) Training time comparison of MRPSF and CNN

5.5.2 准确率 – 速度权衡

本小节简要地评估针对 MRPSF 的准确率 – 速度权衡. 具体而言, 即参数 p 如何影响准确率和时间这两个人们比较关心的指标. 我们从 CASIA-B 数据集随机选取 12 个个体进行实验, 共进行 8 组实验. 由于森林中决策树的数量会明显影响训练时间, 同时也会影响识别准确率, 我们设置不同的森林大小, 记录所需要的训练时间和能够达到的准确率, 并且选取 One-on-one Net 作为深度学习方法的代表, 记录其训练中损失达到基本收敛的时间进行对比. 本实验中, One-on-one Net 没有使用 CUDA 加速技术. 实验结果如图 9 所示.

从图中可以看出, 随着决策树数量提升, 准确率大体呈上升趋势, 这和 5.6.2 小节的结论是一致的. 当决策树数量为 3500 时, MRPSF 的准确率达到最高, 此时训练时间只有深度神经网络的一半左右. 在决策树数量更少, 如 1000~2000 时, 准确率也能达到较好水平. 此时训练时间则比深度神经网络低很多. 本节实验表明, MRPSF 可以在准确率和时间两个指标上提供一定的灵活性.

5.6 算法参数的影响

MRPSF 算法有多个参数, 本小节我们通过实验研究这些参数对算法性能的影响. 我们从 CASIA-B 中随机选取 12 个个体作为一组进行识别实验, 每种状态下都进行 8 组实验, 固定其他参数, 研究某一参数对算法的影响.

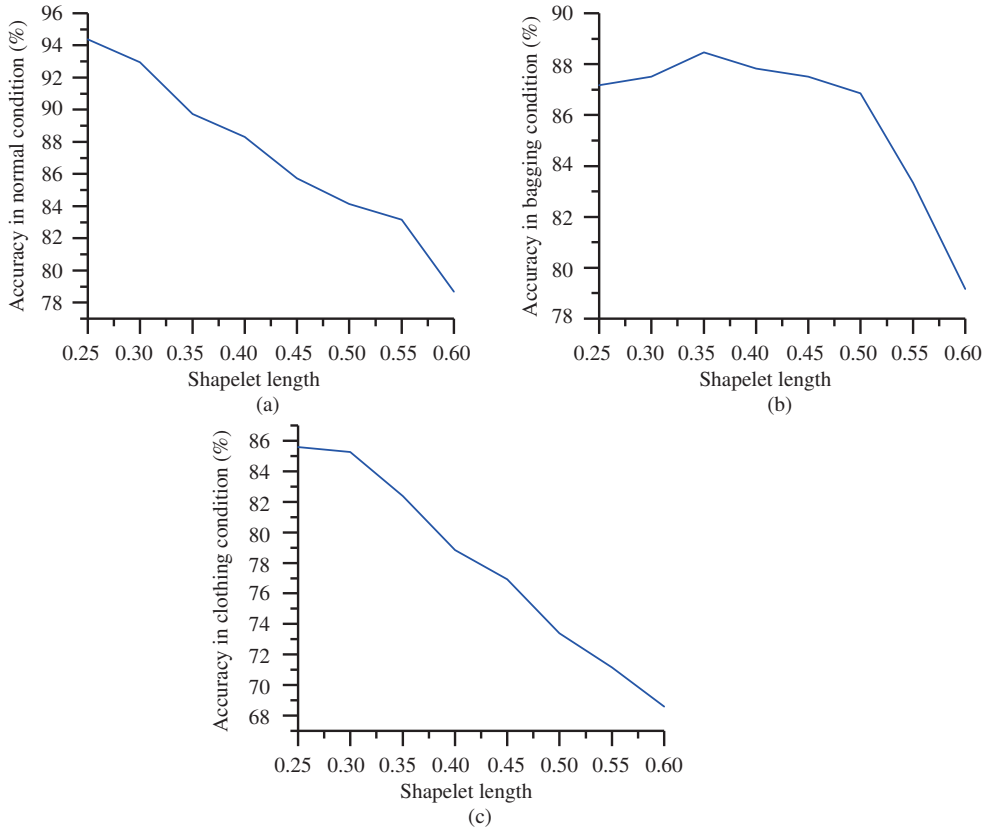

图 10　(网络版彩图) Shapelet 长度对识别准确率的影响

Figure 10　(Color online) Accuracy with Shapelet length increasing. (a) Normal condition; (b) bagging condition; (c) clothing condition

5.6.1　Shapelet 长度上下界 l, u 的影响

本节固定 Shapelet 长度区间的长度为 20% 时间序列长度, 即最大 Shapelet 长度减去最小长度为整条序列长度的 20%, 不断改变该长度窗口的平均值, 观察分类准确率的变化. 注意 20% 的长度区间较小, 这样使得准确率的波动对于该窗口的滑动较为敏感.

实验结果如图 10 所示. 从图中可以看出, 在 Shapelet 长度较小时, 算法的性能表现最好. 这符合步态数据的特点. 由于步态时间序列由多个步态周期构成, 而步态特征往往在一个步态周期之中, 因此特征相对于整条序列的长度较短. 另外我们注意到, 在正常和穿大衣的情况下, 都是最短的 Shapelet 长度区间得到了最好的效果, 在背包的情况下, 则是略长一些的长度效果较好. 实验结果表明, 在使用本文提出的方法进行步态识别时最好使用较短的 Shapelet 长度, 同时可以尝试多个较短的设置, 并且在验证集上测试, 以期得到最佳的效果.

5.6.2　集成 Shapelet 决策树个数 p 的影响

本节考察集成决策树数量 p 的影响. 我们将 p 从 500 递增至 5000, 并记录所有情况的分类准确率. 最小和最大长度比例分别被设置为 0.15 和 0.35. 对于随机森林, 被广泛认可的是, 模型的准确率会随着分类器数量的增加而不断提升.

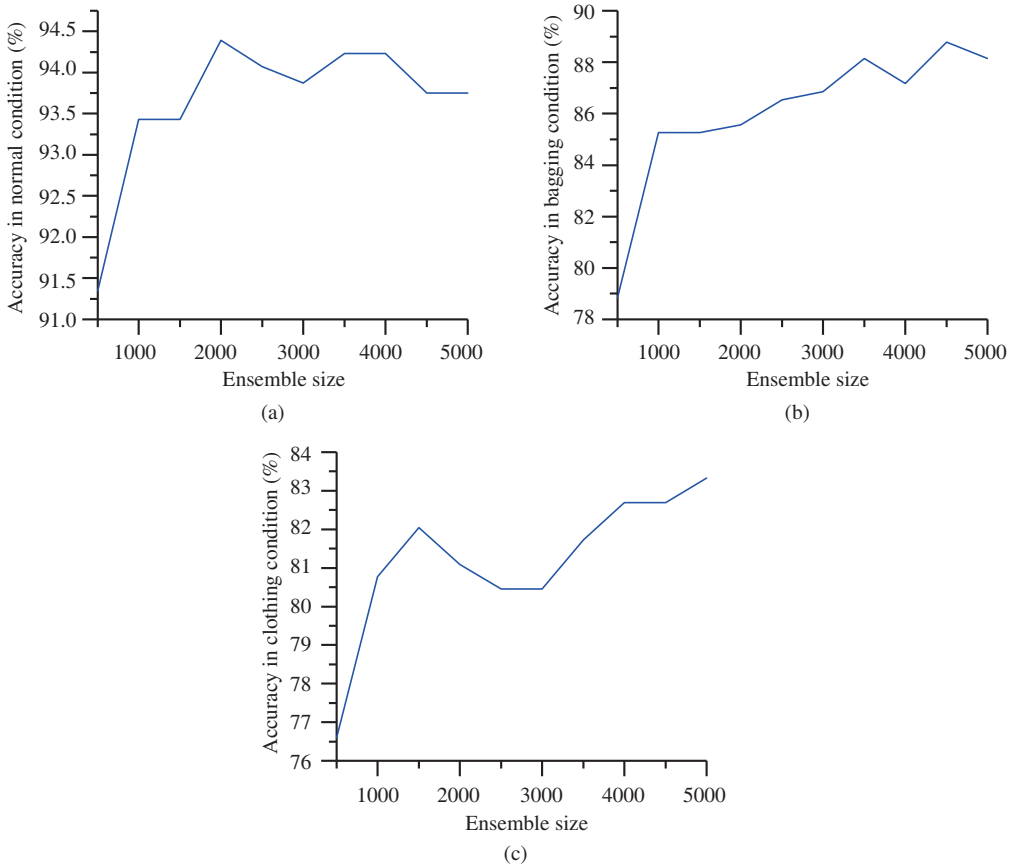

图 11 (网络版彩图) 集成决策树数量对识别准确率的影响

Figure 11 (Color online) Accuracy with ensemble size increasing. (a) Normal condition; (b) bagging condition; (c) clothing condition

图 11 展示了本实验的结果. 观察实验结果可以看出, 在整体趋势上, 较大的森林大小取得了最好的效果, 但不同的情况展现出了不同的特点. 在背包和穿大衣的情况下, 随着森林大小的增加, 识别精确率持续上升. 而在正常行走状态下, 森林大小处于 2000 时就达到了较好的效果, 随着森林大小的继续上升, 准确率在较高处呈现波动. 我们认为, 森林大小 p 的提升在达到某一阈值前有助于准确率的提升, 在该阈值之后, 则不再起作用. 这可以解释实验的结果. 由于正常情况下的识别是比较简单的识别任务 (训练数据较多, 干扰较少), 在该情况下森林大小在 2000 附近就达到了该阈值, 因此在这之后出现了波动. 而背包和穿大衣情况比较困难, 因此准确率呈现不断上升. 然而, 随着 p 的增加, 算法训练和预测消耗的时间也将增加, 因此我们需要权衡准确率和时间之间的矛盾, 尽量在该阈值附近确定 p.

6 结论和展望

本文将步态视频的特征表示为多维时间序列, 将组合 Shapelets 随机森林推广到多维时间序列分类问题中, 并使用该算法进行步态识别, 在小规模问题上取得了能够与目前最先进的方法 —— 基于深度神经网络的方法相接近的准确率. 通过使用 DMDI 森林解释方法, 本文的算法可以细致且详细地解

释模型的决策依据, 从而解决深度神经网络受到诟病的一大问题. 实验表明, 本文提出的方法具有较好的分类准确率, 同时可以提供较好的可解释性, 即可以指出哪些特征在哪几帧对某个个体具有辨别性.

就准确率而言, 本文提出的算法逊于基于深度神经网络的算法. 在这一方面仍有改进的空间. 我们认为原因在于本文方法使用的手工设计的步态特征并非最适合识别任务的特征, 导致 MRPSF 在精度上逊于可以逐层自动抽取特征的深度学习方法. 本文直接使用了在前人工作中被使用的典型步态特征, 并未在特征的选择上进行探索. 后续的工作可以尝试使用更多的时间序列步态特征, 例如和手臂运动相关的特征、轮廓到质心的距离、图像不变矩等, 以期进一步提升识别准确率.

另外, 以上方案对所有步态数据采用同样的步态特征, 并未利用不同训练集的特点. 我们可以借鉴深度神经网络的特性, 参考经典机器学习领域的特征选择方法, 设计数据驱动的自动步态特征选择算法, 以充分学习数据特点, 提升识别准确率.

参考文献

1 Nixon M S, Tan T N, Chellappa R. Human Identification Based on Gait. Boston: Springer, 2005
2 Bouchrika I. A survey of using biometrics for smart visual surveillance: gait recognition. In: Advanced Sciences and Technologies for Security Applications. Berlin: Springer, 2018
3 Han J, Bhanu B. Individual recognition using gait energy image. IEEE Trans Pattern Anal Mach Intell, 2006, 28: 316–322
4 Bashir K, Xiang T, Gong S. Gait recognition using gait entropy image. In: Proceedings of the 3rd International Conference on Imaging for Crime Detection and Prevention, London, 2009. 1–6
5 Bashir K, Xiang T, Gong S. Gait recognition without subject cooperation. Pattern Recogn Lett, 2010, 31: 2052–2060
6 Ji S, Xu W, Yang M, et al. 3D convolutional neural networks for human action recognition. IEEE Trans Pattern Anal Mach Intell, 2013, 35: 221–231
7 Taigman Y, Yang M, Ranzato M A, et al. DeepFace: closing the gap to human-level performance in face verification. In: Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, 2014. 1701–1708
8 Alotaibi M, Mahmood A. Improved gait recognition based on specialized deep convolutional neural network. Comput Vision Image Und, 2017, 164: 103–110
9 Shiraga K, Makihara Y, Muramatsu D, et al. GEINet: view-invariant gait recognition using a convolutional neural network. In: Proceedings of 2016 International Conference on Biometrics, Halmstad, 2016. 1–8
10 Kale A, Sundaresan A, Rajagopalan A N, et al. Identification of humans using gait. IEEE Trans Image Process, 2004, 13: 1163–1173
11 Shajina T, Sivakumar P B. Human gait recognition and classification using time series shapelets. In: Proceedings of 2012 International Conference on Advances in Computing and Communications, Cochin, 2012. 31–34
12 Wang J. Human gait recognition based on one-dimensional motion curve on the side view angle. Dissertation for Master Degree. Beijing: Beijing Jiaotong University, 2017 [王锦. 侧面视角下基于一维运动曲线的人体步态识别. 硕士学位论文. 北京: 北京交通大学, 2017]
13 Yam C Y, Nixon M S, Carter J N. Automated person recognition by walking and running via model-based approaches. Pattern Recogn, 2004, 37: 1057–1072
14 Ye L, Keogh E. Time series shapelets: a new primitive for data mining. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, 2009. 947–956
15 Karlsson I, Papapetrou P, Boström H. Generalized random shapelet forests. Data Min Knowl Disc, 2016, 30: 1053–1085
16 Shi M H, Wang Z H, Yuan J D, et al. Random pairwise shapelets forest. In: Proceedings of the 22nd Pacific-Asia Conference on Knowledge Discovery and Data Mining, Melbourne, 2018. 68–80
17 Liu L, Jia W, Zhu Y. Survey of gait recognition. In: Proceedings of 2009 International Conference on Intelligent Computing, Ulsan, 2009. 652–659
18 Johansson G. Visual perception of biological motion and a model for its analysis. Percept Psychophysics, 1973, 14: 201–211
19 Cunado D, Nixon M S, Carter J N. Using gait as a biometric, via phase-weighted magnitude spectra. In: Proceedings of

International Conference on Audio and Video-Based Biometric Person Authentication, Crans-Montana, 1997. 93–102
20 Fujiyoshi H, Lipton A J, Kanade T. Real-time human motion analysis by image skeletonization. IEICE Trans Inf Syst, 2004, 87: 113–120
21 Rustagi L, Kumar L, Pallai G N. Human gait recognition based on dynamic and static features using generalized regression neural network. In: Proceedings of the 2nd International Conference on Machine Vision, Dubai, 2009. 64–68
22 Iwama H, Okumura M, Makihara Y, et al. The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. IEEE Trans Inform Forensic Secur, 2012, 7: 1511–1521
23 Lishani A O, Boubchir L, Bouridane A. Haralick features for GEI-based human gait recognition. In: Proceedings of the 26th International Conference on Microelectronics, Doha, 2014. 36–39
24 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. Commun ACM, 2017, 60: 84–90
25 Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In: Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, 2015. 1–9
26 Girshick R B, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, 2014. 580–587
27 Girshick R B. Fast R-CNN. In: Proceedings of 2015 IEEE International Conference on Computer Vision, Santiago, 2015. 1440–1448
28 Ren S, He K, Girshick R B, et al. Faster R-CNN: towards real-time object detection with region proposal networks. In: Proceedings of Annual Conference on Neural Information Processing Systems 2015, Montreal, 2015. 91–99
29 He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans Pattern Anal Mach Intell, 2015, 37: 1904–1916
30 Goodfellow I J, Abadie J P, Mirza M, et al. Generative adversarial nets. In: Proceedings of Annual Conference on Neural Information Processing Systems 2014, Montreal, 2014. 2672–2680
31 Isola P, Zhu J, Zhou T, et al. Image-to-image translation with conditional adversarial networks. In: Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 5967–5976
32 Zhu J, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of IEEE International Conference on Computer Vision, Venice, 2017. 2242–2251
33 Feng Y, Li Y, Luo J. Learning effective gait features using LSTM. In: Proceedings of the 23rd International Conference on Pattern Recognition, Cancun, 2016. 325–330
34 Wu Z, Huang Y, Wang L, et al. A comprehensive study on cross-view gait based human identification with deep CNNs. IEEE Trans Pattern Anal Mach Intell, 2017, 39: 209–226
35 Chao H, He Y, Zhang J, et al. GaitSet: regarding gait as a set for cross-view gait recognition. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence, Honolulu, 2019
36 Zhang C, Liu W, Ma H, et al. Siamese neural network based gait recognition for human identification. In: Proceedings of 2016 IEEE International Conference on Acoustics, Speech and Signal Processing, Shanghai, 2016. 2832–2836
37 McLaughlin N, Rincon J M, Miller P C. Recurrent convolutional network for video-based person re-identification. In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016. 1325–1334
38 Yu S, Chen H, Reyes E B G, et al. GaitGAN: invariant gait feature extraction using generative adversarial networks. In: Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, 2017. 532–539
39 Hills J, Lines J, Baranauskas E, et al. Classification of time series by shapelet transformation. Data Min Knowl Disc, 2014, 28: 851–881
40 Yoo J H, Nixon M S, Harris C J. Extracting human gait signatures by body segment properties. In: Proceedings of the 5th IEEE Southwest Symposium on Image Analysis and Interpretation, Santa Fe, 2002. 35–39
41 Breiman L. Random forest. Mach Learn, 2001, 45: 5–32

An interpretable gait recognition method based on time series features

Mohan SHI & Zhihai WANG*

School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044*, China*
* Corresponding author. E-mail: zhhwang@bjtu.edu.cn

Abstract Gait recognition is a type of biometric recognition that can be used as an identification tool in various applications. Deep learning-based methods have recently exhibited promising accuracy in gait recognition tasks; however, in addition to an accurate prediction, these methods are required to explain the recognition results. The black-box nature of deep neural networks makes it very difficult to interpret the basis for their identification. The published studies on the interpretability of gait recognition are also in a blank state. Moreover, deep neural networks require a large amount of data to learn the model parameters and an effective generalization on unseen data is difficult when the problem size is small. Thus, this paper presents a gait recognition method combining accuracy and interpretability. The gait feature is represented as a multi-dimensional time series and a Shapelet-based time series classification method is used for gait recognition. A Shapelet is the most discriminative subsequence in time series that makes the proposed method provide interpretability and accuracy simultaneously. We conducted experiments on the CASIA-B dataset and compared the proposed method with several state-of-the-arts deep learning methods. Experiments show that the proposed method can provide an accuracy close to that of deep neural networks on small-scale data sets. At the same time, the decision-making reason of the model can be explained in detail. Concretely, our method can reveal discriminative gait features and frame numbers for specific subjects.

Keywords gait recognition, time series, Shapelet, random forest, interpretability

Mohan SHI was born in 1993. He received the bachelor degree in computer science and technology from Capital Normal University, Beijing, in 2016. Currently, he is a postgraduate student in Beijing Jiaotong University. His main research interests include machine learning and computer vision.

Zhihai WANG was born in 1963. He received the Ph.D. degree in computer application from Hefei University of Technology, Hefei, in 1998. Currently, he is a professor at Beijing Jiaotong University. His research interests include machine learning and data mining.